13RSL-AFIT1

Workload-Adaptive Human Interface to Aid Robust Decision Making in Human-System Interface

Michael E. Miller, Gilbert Peterson, Brent Langhals, and Jason Bindewald

April 30, 2014

Year 1 Report, April 2013 – April 2014

Workload-Adaptive Human Interface to Aid Robust Decision Making in Human-System Interface

Principal Investigator: Michael E. Miller

Air Force Institute of Technology

Email: michael.miller@afit.edu; Phone: (937) 352-6122 x4651

Research Objectives: The project objective is to demonstrate a system with selective automation, which estimates human mental workload and modulates the use of automation to maintain the mental workload of the operator within a desired range. This system will estimate workload from projected task load, modifying the interface to provide a stable workload. The workload estimate will be updated based upon physiology feedback from the operator. The demonstrator will ideally include 1) demonstration of an interactive workload-adaptive interface; 2) Tuned learning algorithms that can learn the user's workload response and automation preferences with minimal operator interaction; and 3) A robust method for estimating human mental workload from system generated tasks.

Technical Summary: First year focused primarily on designing the interactive workload adaptive interface. To this end, an initial design method was developed to formalize the design of the adaptive interface. Additionally, an initial experiment was conducted which indicated the ability to learn user behavior when interacting with the system, which inspired a concept for adapting the workload adaptive interface based upon user behavior. This line of research is being continued through an ongoing experiment. Finally, a pair of experiments were conducted which provided more insight into the use of physiology metrics for tuning the workload-adaptive interface. This work resulted in one journal article, in review, and three conference papers, one of which has recently been awarded Best Human Factors Paper at the Industrial and Systems Engineering Research Conference.

Our research has already provided a new adaptive automation design model, allowing system designers the ability to visually and systematically evaluate the placement of adaptive automation within a system network. Further, this method can help the designer to isolate the tasks, which require human decision making, ideally permitting

tasks to be automated which do not require human decisions. The Space Navigator platform designed using the model, has allowed us (and future researchers) to simplify data gathering from human participants. The resulting automation system will allow for research into how similarity and difference of actions between a human-machine team affect the overall performance of the system. The final experimental data will provide several areas of further research including trust in automation, training improvement, workload reduction (actual and perceived), and task load switching.

Design Framework

An interactive environment was created which was derived from an open source routing game (Beebe and Beebe, 2010). To facilitate the design, it was necessary to develop a more comprehensive method for the design of adaptive automation systems. An overview of the process is shown in Figure 1. An original description of this process is provided in Appendix C. As shown, the design process begins with establishing the overall goal(s) of the system and decomposing these goals into functions and subfunctions until each sub-function can be reasonably allocated to a man or machine (Steps 1 through 3), as is common in man-machine interface design. Next, a function relationship diagram is developed as indicated in Step 4. This function relationship diagram clearly depicts the interdependencies among and the sequencing of the leaf-level functions from the function decomposition. Note that in this process, we clearly differentiate the terms "function" from "task" with function referring to unallocated processes that must be accomplished to fulfill the goal and "task" referring to the steps which must be performed by an allocated entity (e.g., man or machine) to accomplish the functions.

The framework, relies upon a revised version of a task allocation diagram originally devised by Price (Price, 1985) to create an initial allocation of tasks to a man or machine as shown in Figure 2. This figure permits classification of functions into tasks that can be performed best by man, machine, or as tasks that can be performed nearly equally by either entity. With this in mind, tasks are initially allocated to the more capable entity, step 5 in the process of Figure 1. However, tasks that can performed equally by either entity are originally classified to be adaptively allocated. A task relationship diagram is then developed from the function relationship diagram by replicating the functions as tasks, but applying color and patterns for each task, which code the initial allocation to create an initial task relationship diagram. The framework further acknowledges that edges between two tasks within the task relationship diagram, where one task is allocated to a man and the other to a machine, include a pair of inherent tasks. These inherent tasks arise as one entity must convey information, which the other entity must receive. These inherent tasks are also depicted within the task relationship diagram.
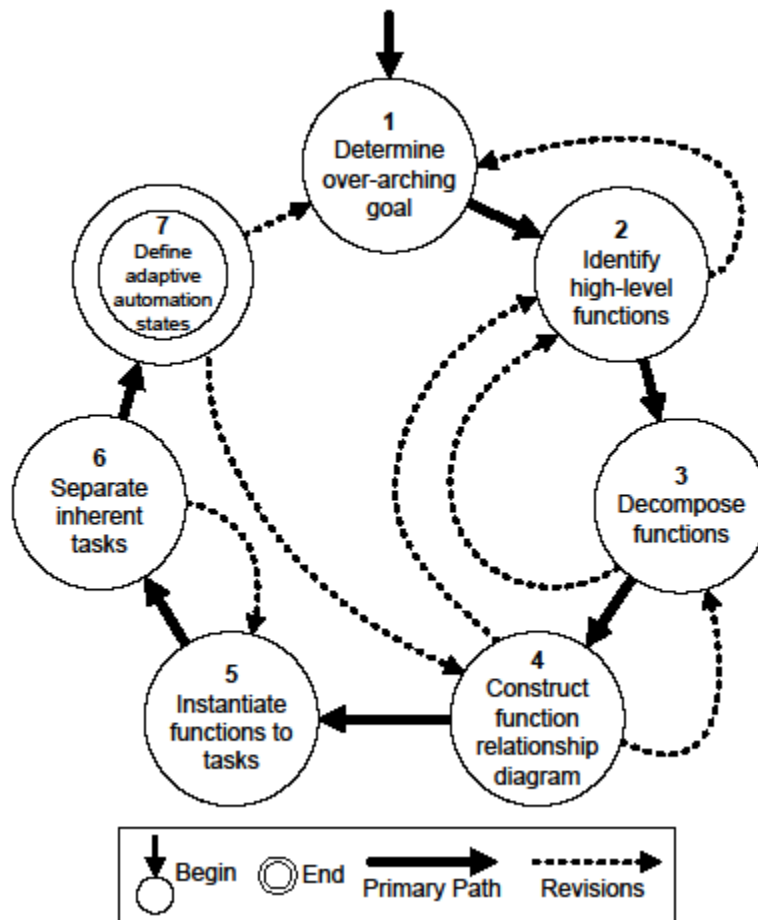
Figure 1. Automation Design Framework Process Diagram

The resulting task relationship diagram then permits the designer to visualize a number of important design tradeoffs. Specifically, the number tasks to be adapted will be shown through color-coding and the number of disconnected adaptive nodes indicates the number of unique adaptive states, which the system must be designed to accommodate. Therefore, large numbers of disconnected adaptive nodes indicate a large number of unique states and added system complexity. Highly interconnected groups of functions can be clustered into groups. The allocation of these clusters to a single entity also helps to reduce the complexity and the interdependence of the entities. The task handoffs between any two entities are also clearly indicated and handoffs, which require the exchange of complex information, can further be coded to help the designer understand that these exchanges should be considered for elimination. Areas of the diagram, which contain single connections between clusters of nodes, indicate groups of tasks that have few dependencies and therefore provide an advantageous location for a handoff between a man and a machine. Further, the number of inherent tasks can be high for a given design, prompting a redesign of the

resulting diagram. Through this visualization, the designer is able to loop among the process steps shown in Figure 1, to derive a desirable task relationship diagram.
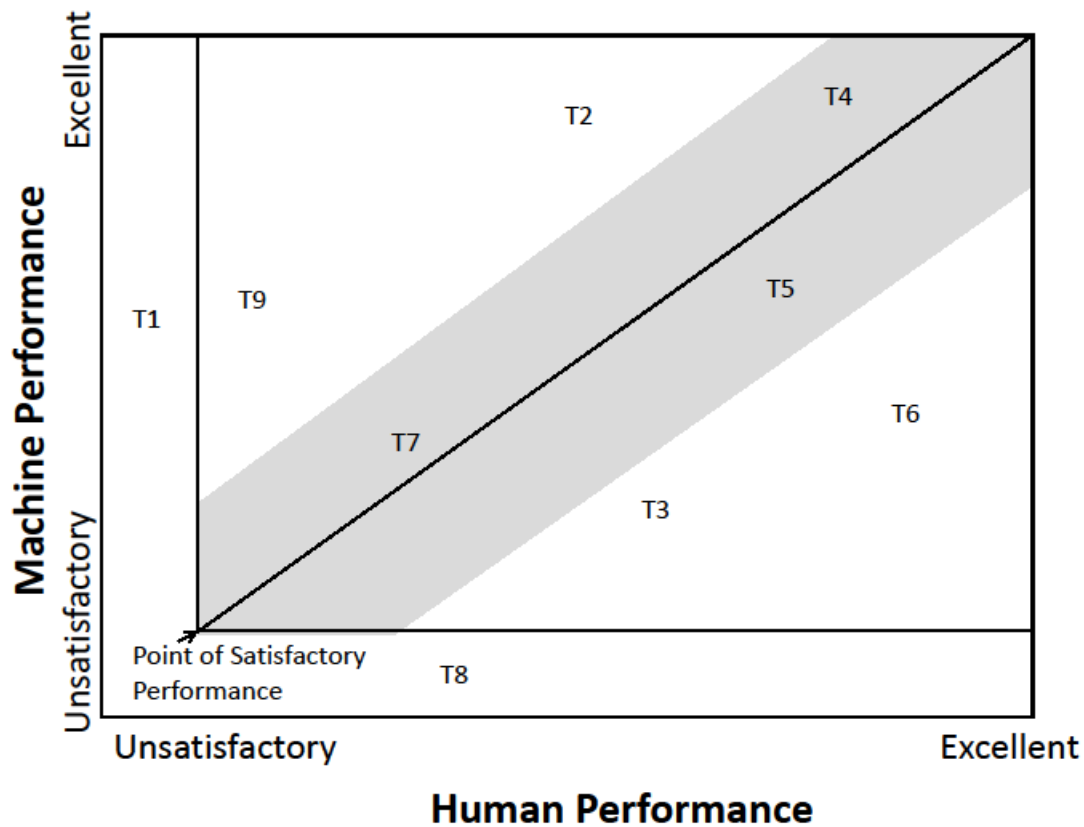

Figure 2. Depiction of Tasks (T1-T9) based upon the ability of a man or machine to complete the task. Note the shaded region indicate tasks desirable for adaptive allocation as they can be performed nearly equally well by man and machine.

Figure 3 shows an example task relationship diagram created during the design of the adaptive interface. Inherent tasks, which stem from the need for communication between the human and machine, are explicitly shown through the square C/P nodes. Allocation is shown through hash marks within the task nodes. Desired allocation based upon the initial allocation, while not shown in this monochrome figure, could be indicated through color coding. It is notable that operator workload is not explicitly captured at this time within the task relationship diagram. Research to be conducted during the coming year will extend this method to further include human workload within the allocation process. An initial experiment and workload model development using the US Army Research Laboratory's Improved Performance Research Integration Tool are currently underway.
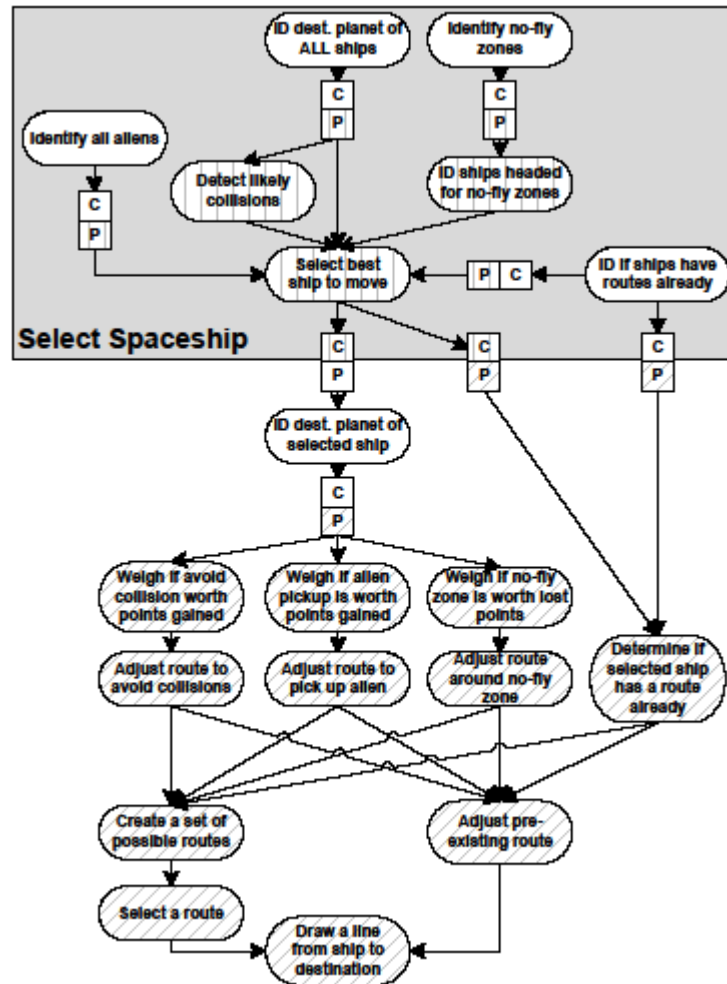
Figure 3. Example Task Relationship Diagram. Nodes containing diagonal hash marks are allocated to the user, vertical hash marks indicate adaptive nodes. C/P nodes represent inherent tasks where information must be communicated and perceived as control is passed from the human to machine or vice versa.

Learning Algorithms

We created an environment for investigating adaptive automation. The resulting system is a tablet-based game called Space Navigator where the operator draws trajectories on the screen to interact with the system. The Space Navigator environment resembles a multi-UAV routing operation. In which UAVs are tasked to the operator, and the operator must route the UAVS to a specified destinations without colliding with other UAVs

and avoiding danger zones. The Space Navigator environment is an open-source system that can be easily ported to different types of devices that includes several data gathering for user responses, and the NASA TLX to measure workload. A secondary goal of the designed system was to simplify the data collection process by "gamifying"

the system (Hamari, Koivisto, and Sarsa 2014). By being a game and open source, it provides a means for others to extend the work we are doing.

A similarity measure similar to Modified Hausdorff Distance (Atev, Masoud, and Papanikolopoulos 2006), called Windowed Hausdorff Distance, has been developed in order to compare trajectories of different lengths. Coupling this similarity measure with the player profiling system will form the base of an adaptive automation system to play Space Navigator.

A recently completed data-collection experiment has facilitated the creation of a game-play database. Preliminary results confirm that a player-by-player discernment is not feasible, but that player profiles can be created to discern between larger groups of players with specific tendencies. An unsupervised learning system will be used to cluster users. The learning will use an agglomerative clustering algorithm, and a Fisher score supervised feature selection algorithm to determine what defines different clusters of users. The efforts to complete the player profiling system will be completed over the summer of 2014.

The next step in determining how similarity of action affects human-machine system performance is to create a computational system to mimic human game-play patterns. The objective of this study is to see to what extent we can distinguish between specific players of the game. Presently, we are working to take the game-play database and use it to create a player profiling system.

The resultant system and experiments based on it will allow us to answer the question: "How does the similarity or dissimilarity of the automated aid's task performance to that of the operator affect the overall human-machine team's performance?"

Workload Estimation and Physiology Indication

Research was conducted to develop physiology measures, which could be used to estimate workload for individuals and ideally adjust workload estimates as a function of task load. We conducted a pair of studies towards' this end.

In a first study, described in detail in Appendix D, electrocardiography measures were recorded while each of 13 participants performed tasks using MAT-B with one of four task loads, ranging from a medium task load to a high task load (Splawn and Miller, 2013). The high task load within this experiment was purposefully design to present the operator with conditions, which were practically impossible for participants to complete with an exceptional level of performance. Various heart-related metrics, including heart rate, high and low frequency heart rate variability (hfhrv, lfhrv), standard deviation of NN intervals (SDNN) and the coefficient of variation of the R-R interval (CVRR). Additionally, the difference in heart rate and hfhrv with a baseline value were also

7

calculated. This research indicated that the hfhrv, SDNN, heart rate delta from baseline and the hfhrv delta from baseline are correlated with task load. Additionally, a linear model of performance in terms of response time, heart rate delta, hfhrv delta and the hfhrv accounted for 55% of the variance in NASA TLX ratings of perceived workload. This model indicated that a combination of appropriately selected heart rate measures with response time measures could inform a system as to the relative workload of a user.

In a second study, described in detail in Appendix E, participants were exposed to a 40 minute, difficult vigilance task (Jeroski, et al, 2014). In addition to electrocardiography, cerebral oximetry and electrooculography were performed. Heart rate measures, cerebral oxygen values and blink rates were correlated with performance. CVRR was shown to be significantly correlated with task performance and cerebral oxygen level was negatively correlated with task performance. Analysis of this data is ongoing and additional physiology metrics will be explored.

Together these two studies demonstrate the use of physiology metrics to gauge workload and performance. Results from these two studies will be used to formulate a real-time physiology metric to be incorporated into the final system design.

Funding Summary by Cost Category – Be Specific (by FY, $K):

Funding Utilized ($K)	Starting FY (Actual)	FY+1 (Estimated)	FY+2 (Estimated)
Non-Military Government Personnel Costs	$17,240.61	$30,670	$31,591
In-house Contractor Costs	$0.00		
Travel (Be Specific)	$0.00	$5000	$5000
Training (Be Specific)			
Supplies	$3480.67	$2500	$2500
Other Expenses (Be Specific)	Equipment $5669.12 Incr. Cost $1657.70	Equipment $2500 Incr. Cost $2290	Equipment $2500 Incr. Cost $2345
Total Resource Requirements	$28048.10	$42,960	$43,936

References

Atev, S.; Masoud, O.; and Papanikolopoulos, N. 2006. Learning traffic patterns at intersections by spectral clustering of motion trajectories. In Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on, 4851-4856.

J. Beebe, B. Beebe, Flight control clone with corona in 30 hours (December 2010). URL: http://jonbeebe.tumblr.com/post/2301674980/flight-control-clone-with-corona-in-30-hours

Hamari, J.; Koivisto, J.; and Sarsa, H. 2014. Does gamification work? A literature review of empirical studies on gamification. In Proceedings of the 47th Hawaii International Conference on System Sciences. HICSS.

Jeroski, J., Miller, M.E., Langhals, B. and Tripp, L. (in press). Impact of Vigilance Decrement upon Physiology Measures. Proceedings of the Industrial and Systems Engineering Research Conference, Montreal, CA.

Price, H. E., 1985. The allocation of functions in systems. *Human Factors, 27 (1),* 33-45.

Splawn, J. and Miller M.E. (2013) Prediction of perceived workload from task performance and heart rate measures, Proceedings of the Human Factors and Ergonomics Society, San Diego, CA.

Appendix A: In House Activities

Personnel:

Name	Degree	Discipline	Involvement
Air Force Employees:			
Michael Miller	PhD	Sys Engr., Human Factors	(1/12)
Gilbert Peterson	PhD		
LtCol Brent Langhalls	PhD	IT Systems	
Capt. Jason Bindewald	M.S.		
Justine Jeroski	M.S.	Sys Engr, Human Factors	(3/4)
Lt. Mark Harris	B.S.	Sys Engr, Human Factors	(1/10)

On-site Contractors:
   N/A

Visitors:
   N/A

Publications:

Published in Peer Reviewed Journals, Books, etc:

N/A

Published in Un-reviewed Literature (e.g., Technical Proceedings):

Splawn, J. and Miller M.E. (2013) Prediction of perceived workload from task performance and heart rate measures, Proceedings of the Human Factors and Ergonomics Society, San Diego, CA.

Accepted/Submitted for Publication:

Bindewald, J.M., Miller, M.E. and Peterson, G.L. (in review). A method for rationalizing allocation for systems employing adaptive automation. *International Journal of Human-Computer Studies.*

Jeroski, J., Miller, M.E., Langhals, B. and Tripp, L. (in press). Impact of Vigilance Decrement upon Physiology Measures. Proceedings of the Industrial and Systems Engineering Research Conference, Montreal, CA. (Awarded Best Paper - Human Factors).

Invention Disclosures and Patents Granted:

N/A

Invited Lectures, Presentations, Talks, etc:

N/A

Professional Activities (editorships, conference & society committees, etc):

Chair Special Session on Adaptive Automation, Industrial and Systems Engineering Research Conference, Montreal, CA.

Paper Review Committee, Industrial and Systems Engineering Research Conference, Montreal, CA

Paper Review Committee, Human Factors and Ergonomics Conference, Chicago, IL

Honors Received (include lifetime honors such as Fellow, Honorary, Doctorates, etc—also state year elected:

N/A

Extended Scientific Visits From and To Other Laboratories:

N/A

Appendix B: Technology Assists, Transitions, or Transfers—Detailed Listing

Task Title	Performance Period	AFOSR Program Manager	TD Performer (PI and org)	Customer (name, organization, email, & phone)
N/A				

Table Contd.

Research Result (scientific statement)	Application (technical benefits and/or customer use; include and underline military applications first)	From	To	Application

Transitioned From:
AFRL=L; Industry = I; Academia = A

Transition To:
Industry = I
Air Force 6.2 or 6.3 = AF;
Other AF, DOD, Government, etc. = O

Application
Product (New or Improved) = Pd;
Process (New or Improved) = Pc;
Other = O (please specify)

Appendix C

Submitted Article: Bindewald, J.M., Miller, M.E. and Peterson, G.L. (in review). A method for rationalizing allocation for systems employing adaptive automation. *International Journal of Human-Computer Studies.*

A Function-to-Task Process Model for Adaptive Automation System Design

Jason M. Bindewald[a,*], Michael E. Miller[b], Gilbert L. Peterson[a]

[a]*Department of Electrical and Computer Engineering, Air Force Institute of Technology, 1950 Hobson Way, Wright-Patterson AFB, OH 45433-7765*
[b]*Department of Systems and Engineering Management, Air Force Institute of Technology, 1950 Hobson Way, Wright-Patterson AFB, OH 45433-7765*

Abstract

Systems have been proposed in which the level of automation adapts in real-time to maintain engagement of the human operator while preventing operator overload. Unlike traditional systems that allocate functions to either the human or the machine, adaptive automation varies the allocation of functions during system operation. Designing these systems requires designers to consider issues not present during static system development. To assist in adaptive automation system design, this paper presents a design process model for determining how to allocate functions to the human, machine, or dynamically between the two. An illustration of the process demonstrates the potential complexity inherent in adaptive automation systems and how the process model aids in understanding this complexity during early stage design.

Keywords: Adaptive Automation, Function Allocation, Man-Machine Allocation, Network, Interaction Design

1. Background

Consumer, commercial, and government systems increasingly apply automation, particularly in systems which involve time critical decisions and actions. These systems include manufacturing plant process control (Itoh et al., 1999; Valente et al., 2010; Valente and Carpanzano, 2011), aircrew and air traffic control (Prevot et al., 2008), and remotely piloted or controlled vehicles (Parasuraman and Wickens, 2008; Parasuraman et al., 2009; Kidwell et al., 2012). Automation can improve the performance of systems without increasing manpower requirements by allocating routine tasks to automated aids, improving safety through the use of automated monitoring aids, and reducing the overall

---

*Corresponding author. Tel.: +1 937 255 3636 x6129; fax: +1 937 656 7061
Email addresses: Jason.Bindewald@afit.edu (Jason M. Bindewald),
Michael.Miller@afit.edu (Michael E. Miller), Gilbert.Peterson@afit.edu (Gilbert L. Peterson)

cost or improving productivity of systems (Rouse, 1981). Additionally, automation can permit removal of the operator from particularly undesirable or dangerous environments (Nakazawa, 1993), increasing the safety and reducing stressors placed upon the operator.

Unfortunately, automation system designers have limited ability to project future events, and are often unable to adapt when unforeseen circumstances occur. As such, utilization of a human operator who can adapt to these unforeseen circumstances to provide system resilience is desirable (Woods and Cook, 2006). With the inclusion of a human operator, other problems often arise. Some include operator over-reliance on automation (Itoh, 2011), operators placing inappropriate levels of trust in the automation (Dzindolet et al., 2003; Lee and See, 2004; Merritt et al., 2012), or operators losing situation awareness to preclude appropriate recovery from automation failures (Itoh, 2011). Further, as operators are not performing active control of the system, they may not practice the knowledge necessary to operate the system and can suffer from skill atrophy (Kirwan, 2005). As a result, adaptive automation systems have been proposed to maintain user engagement, without overloading operators (Rouse, 1977).

Automation is the capability "to have a computer carry out certain functions that the human operator would normally perform" (Parasuraman et al., 2000). Which entity performs a given task helps determine whether to automate a task or not. There are many types of tasks, and consequently, several forms of automation. The categories of automation can include, "the mechanization and integration of the sensing of environmental variables; data processing and decision making; mechanical action; and 'information action' by communication of processed information to people" (Sheridan and Parasuraman, 2005).

According to Merriam-Webster's Dictionary, the definition of *adapt* is "to make fit (as for a new use) often by modification" (Merriam-Webster, 2013). In this regard, adaptation can be useful in helping to move between the different types of automation, and between different levels and stages of automation. Since a dynamic approach to automated decision-making was proposed by Rouse (1981, 1977), the field has adopted the terms *adaptive automation* and *adaptive systems* to define the idea of an automated system that can adapt to a changing environment.

Within published research, the definition of adaptive automation has been subject to debate. Most authors would agree that levels or types of automation change in an adaptive system. For example, Dorneich et al. (2012) define adaptive systems as those "allowing the system to invoke varying levels of automation support in real time during task execution, often on the basis of its assessment of the current context...invoking them only as needed". This view of adaptive automation places the onus of determining the current automation state on the system. However, others have shown that even the determination of who 'adapts' the system (e.g., the system, the operator, etc.) can fall on a sliding scale (Parasuraman and Wickens, 2008).

Within the current context, a *system* is a combination of hardware, software, and human operators that work together to accomplish one or more goals. As the

2

focus of the paper is system design, the term *machine* refers to the combination of all hardware and software within the system with which the human operator interacts. The term *task load* describes the number and difficulty of tasks assigned to human operators, to which they must respond. These tasks can be explicitly or implicitly imposed within the operational context of the system–the system might require the operator to make a selection, requiring an explicit action. However, to make this selection, the operator will need to gather appropriate information from the system or environment and make decisions, each of which are implicit task demands.

The term *workload* refers to the perceived impact of the task demand placed upon the operator's mental or physical resources and workload corresponds to the utilization of these resources. The operator will perceive the effect on workload level, and therefore, *perceived workload* corresponds to the user's perception of the degree to which their mental or physical resources are fully utilized. The variability in the task load imposed upon an operator–and the workload the operator experiences–originates from a number of sources. In addition to the variance of performance due to explicitly defined workload, the performance of the human operator–and therefore the impact of a given task load upon their perceived workload–may vary due to inherent factors such as fatigue, stress level, motivation, and training level (MacDonald, 2003; Reid and Nygren, 1988). As such, the ability of a human operator to respond to tasks imposed upon them varies over time (Colombi et al., 2012).

This research investigates the impact of explicit and inherent task load to provide a process model for the design of adaptive automation systems. The resulting function-to-task design process model creates a set of visual diagrams enabling designers to better allocate tasks between human and machine. This is achieved through a set of five analysis tools allowing designers to identify points within a function network where the transitions between human and machine entities can facilitate adaptive automation. This paper proceeds as follows. Section 2 reviews the design processes currently in place for adaptive automation systems. Section 3 presents the function-to-task design process model. Section 4 illustrates the function-to-task design process model through a system design iteration.

2. Designing Adaptive Automation Systems

Discussions on the design of manned systems as a tool to aid allocation of functions or tasks between a human operator and a machine often cite Fitts's list (Fitts et al., 1951). Fitts *et al* discussed tasking the machine to perform routine tasks that require high speed and force, computational power, short-term storage, or simultaneous activities. Fitts *et al* further propose leveraging the human's flexibility, judgment, selective recall, and inductive reasoning to improve system robustness to unforeseen circumstances. Fitts *et al* also acknowledge the limitation of humans to correctly employ these capabilities when overloaded due to excessive task demands or to maintain alertness and employ these capabilities when not actively participating in system control.

One may consider the allocation of functions between man and machine within a system as a multi-objective optimization, wherein designers optimize some combination of performance, safety, and robustness as a function of the tasks allocated to each component. The limitations of system and human capability shape this optimization, with a significant component of human capability quantified in terms of human workload. Adaptive automation system design assumes that the number and difficulty of tasks performed will vary over time, and the tasks allocated to the human or machine need to vary to provide the human operator with an appropriate workload.

Figure 1 illustrates this concept, which depicts a two-dimensional space which arranges tasks, T1-T9, based on how well a human operator or the machine can perform them under reasonable task load. As shown, performance by either system can range from unsatisfactory through excellent (Price, 1985). We should allocate tasks, such as T1 or T8–which one entity (human or machine) can perform more satisfactorily, than the other entity–to the better performing entity. However, any task that either entity can perform beyond the point of satisfactory performance, we can reasonably allocate to either human or machine.

If there was no constraint on resources, one could then maximize performance of the overall system by allocating tasks below the 45 degree line to the human and tasks above this line to the machine. However, resource constraints force a shift in the location of this line. For instance, assuming workload limits on human performance and unbounded machine resources might induce the designer to shift the dividing line lower in the plot, decreasing human workload and allocating additional tasks to the machine. On the other hand, if users' performances improve by increasing their engagement with the system, one may wish to raise the dividing line and allocate more tasks to the human. Therefore, adaptive automation effectively requires the system to permit this allocation line to shift up and down within this plot, allocating fewer or greater numbers of tasks to the human operator. Of course, this strategy assumes that bandwidth does not constrain the machine.
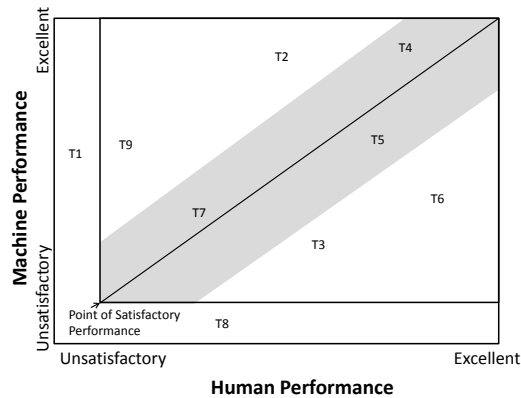

Figure 1: Task Allocation in Adaptive Automation (Adapted from (Price, 1985)).

2.1. Automation Taxonomies

Taxonomies for adaptive automation have been proposed to accommodate the complex design space present in many of the systems. Feigh et al. (2012) indicate that modifying the allocation of tasks among humans or machines can affect operator workload. However, modification of task scheduling, interaction required between the operator and other system elements, or the content of any interaction can also affect operator workload. Although not explicitly captured, these modifications may involve systems with multiple machines or multiple humans (Calleja and Troost, 2005).

Considering the interaction between an individual operator and a machine, Parasuraman et al. (2000) proposed a model for describing levels of automation that builds upon the work of Sheridan and Verplank (1978) to discern between the types and levels of automation. The model delineates the types of tasks performed based on the four-stages of human information processing: sensory processing, perception/working memory, decision making, and response selection. Within these four stages, they take the ideas Sheridan and Verplank propose and codify them further into a 10-point scale describing the levels of automation, ranging from "1. The computer offers no assistance; human must take all decisions and actions" all the way to "10. The computer decides everything, acts autonomously, ignoring the human."

Alternatively, Endsley (1999) propose four core human functions that a system could automate independently of one another, including: monitoring, generating alternatives, selecting alternatives, and implementing the selected alternative. This framework assigns each of these four tasks to either the human or machine (or both in some cases) and enumerates the level of automation between fully autonomous and fully human-implemented, providing a two-dimensional space over which to define automation. Each of these classification schemes permits the differentiation between intermediate levels of automation, explicitly defining which human task a given level automates. Each model aids the creation and classification of automation states for tasks the human or machine can perform, helping the system designer determine "what" to automate and "to what extent"(i.e. level of automation). Although designers can apply "level of automation" models to any system employing automation, they are important in systems employing adaptive automation as they permit the designer to determine what part of and how to automate a task so that changes in automation level can be clearly described.

Although the adaptive automation taxonomy Feigh et al. (2012) propose does not fully overlap the automation taxonomies provided by either Parasuram Parasuraman et al. (2000) or Endsley (1999), the taxonomies are not independent of one another. Feigh *et al* uniquely highlight the fact that not all tasks are time critical, and systems can reprioritize them during periods of peak workload. They also discuss the allocation of tasks between humans and machines– alluding to various levels for automation of tasks that include selection or implementation of alternatives. Additionally, they contend that automation of generating alternatives and monitoring requires automatically generated

information displayed to the human operator, forcing a change in the interaction and content of interaction. Each of these methods, therefore provides a different way to classify and consider the effect of changes in autonomy on operator workload.

2.2. Trigger Taxonomies

In adaptive automation systems, the designer not only decides which tasks to automate, they must decide how to trigger or initiate changes in automation. A common source of human error occurs when an operator assumes that a system is operating in a different mode than it is truly operating. It has been suggested that the designer give complete flexibility and control over task automation determination to the human supervisor (Parasuraman and Wickens, 2008). Unfortunately, in a time constrained environment, making and indicating this decision to the system requires time and increases operator workload. Therefore, considering other possible triggers for changes in automation provides value.

Feigh et al. (2012) proposes five general classes of potential triggers, including operator-based: based on some overt (e.g., human selection) or covert (e.g., a rise in heart rate, indicating a stress response) human input; system-based: based on the current system state (e.g., the presence of an excessive queue of tasks awaiting user response); environment-based: based on the current state of the environment or a specific event; task- and mission-based: based on the initialization/completion of milestones; and spatiotemporal: based on time or location of the system. As a number of potential triggers exist, the designer must select from among a large number of triggers to determine when to change the automation level. To increase the level of automation, the designer may apply different triggers or levels of triggers than to decrease the level of automation.

2.3. Human-Machine Interaction

The need to provide effective communication between the human and machine impedes human interaction with automated systems. In some cases–such as flight control automation–the design of this interaction can have life-or-death consequences (Geiselman et al., 2013; Kaber et al., 2001). Unfortunately, this interaction can become increasingly complex in systems employing adaptive automation. William Rouse's analysis of human-machine interaction within a dynamic system is a seminal article in this field (Rouse, 1981). Rouse shows the different forms of communication with the system as a set of five interaction loops. The first two loops, in which it is possible that no communication is required, represented manual control and completely automated control. In the third loop, wherein he coins the term *overt communication*, the human and machine operators of a system directly communicate information about their tasks. The human operator must take explicit actions to control the machine, and the machine must explicitly provide information. The human operator must consciously read, listen to, or otherwise receive this information. The last two loops represent more subtle communication which typically occurs among humans; *covert communication*, with the fourth loop representing covert human to machine communication and

the fifth covert machine-to-human communication. Information communicated indirectly–which might include state information–characterizes covert communication. The timeliness of a response from a teammate, where hesitancy in response signals uncertainty and fast authoritative response indicates certainty, provides an example of covert communication.

Unfortunately, communication errors occur between human operators and machines as the machine can fail to communicate critical state information (Geiselman et al., 2013), let alone the information leading to the selection of a critical state, or less direct information, such as the certainty of this information. Recent research focuses on improving covert communication from the human to the machine through the use of psychophysiological measures, such as electroencephalography (EEG), electrocardiography (ECG) , electrodermal activity (EDA), electromyography (EMG) (Dorneich et al., 2012; Byrne and Parasuraman, 1996; Haarmann et al., 2009) or behavioral measures, such as eye gaze patterns. Such measures have the "potential to yield real-time estimates of mental state" (Byrne and Parasuraman, 1996), thus allowing the machine to gain information regarding the state of the human operator.

The infeasibility of communicating all automated tasks from a machine to a human aside, the human in an automated system requires enough information to permit appropriate situational awareness. Since the human operator assumes control in the event of a mishap or in order to make a critical decision, the human needs an understanding of the current system and environment state. Several research efforts devote effort toward finding an appropriate balance between providing enough information for situation awareness and overloading the human operator with information (Wickens, 2008; Endsley, 1999; Kaber et al., 2001; Sheridan and Parasuraman, 2005; Manzey et al., 2012; Parasuraman et al., 2008). Systems can present some information more effectively using visual, auditory, tactile or other human senses. Further, all communication will affect the user's workload, potentially resulting in overload conditions. However, the relationship between how humans attend to, receive, process, and act upon information creates complexity, and the interaction influences the human operator's perceived workload (Wickens, 2008).

The types of feedback given influence the resulting system. For example, Manzey *et al* demonstrate that users are much more likely to develop a proper level of trust with a system when the system gives them negative feedback loops rather than positive ones (Manzey et al., 2012). Further issues such as how to design a system to manage interruptions in a socially acceptable manner, and analyzing the positive and negative consequences of automating the interruption management task (Dorneich et al., 2012) are important. The idea of etiquette flows naturally into the concept of trust, directly impacting the human operator's trust of the system.

While the design of the human-machine interface can be complex, this interface requires grounding in an understanding of the information that the human operator and the machine must communicate to facilitate task completion. The importance of this information necessitates its presentation in a way that does not overload the operator and recognizes the fact that the human operator

will not necessarily receive all information the system provides.

3. Function-to-Task Design Process Model

We now explain how the previous adaptive automation design techniques can be augmented to emphasize explicit and inherent task load. The function-to-task process model for the design of adaptive automation systems is presented here to create a set of visual diagrams enabling designers to better allocate tasks between human and machine. The goal of this process is to enable designers by providing a set of analysis tools to identify points within a function network where the transitions between human and machine entities are advantageous for adaptive automation.

The literature recognizes that most functions are made up of sub-functions that are completed in a temporal sequence (Ross, 1977). Although the terms *function* and *task* are sometimes applied interchangeably (Bye et al., 1999) and multiple definitions exist (Concepts and Group, 1998), clear differentiation of these terms leads to a better understanding of the proposed process model. Here, we define a *function* as an action that an element or elements of a system performs to accomplish the desired goals or to provide the desired capability. A function is delineated from a task as the function is not allocated to an entity. A *task* is allocated to a specific entity and represents the actions necessary for the entity to perform the function. Tasks can be explicit, in that the function indicates them; or implicit, in that they are not required by the function but are necessary to enable the entity to perform the function.

The literature on adaptive automation relies on different types or levels of automation. However, adaptive automation levels are often spoken of in continuous terms, which can be thought of metaphorically as a radio volume dial, that increases or decreases an "amount" of automation. However, in the proposed model, the channel selection dial provides a more fitting metaphor as it permits the selection of discrete states. Each function consists of several atomic functions. *Atomic functions* are functions that can only be performed by a single cognitive entity. Further decomposition of an atomic function is not possible, making the determination of automation state a discrete decision. A channel dial serves as a better analogy than a volume dial as each automation state consists of some set of atomic functions.

With these definitions, we now turn our attention to a proposed process model for allocation of functions to entities (e.g., human or machine) and a progression to a task relationship diagram. This process model allows a designer to make informed decisions relating to automation design, particularly in the case of adaptive automation. Figure 2 graphically depicts the function-to-task design process model. The function-to-task process usually proceeds in a linear fashion, as indicated by the solid bold arrows in Figure 2. In some cases, completing steps in the process will force the design back to a previous step for revisions; likely locations for revision steps are indicated by the dashed arrows in Figure 2.
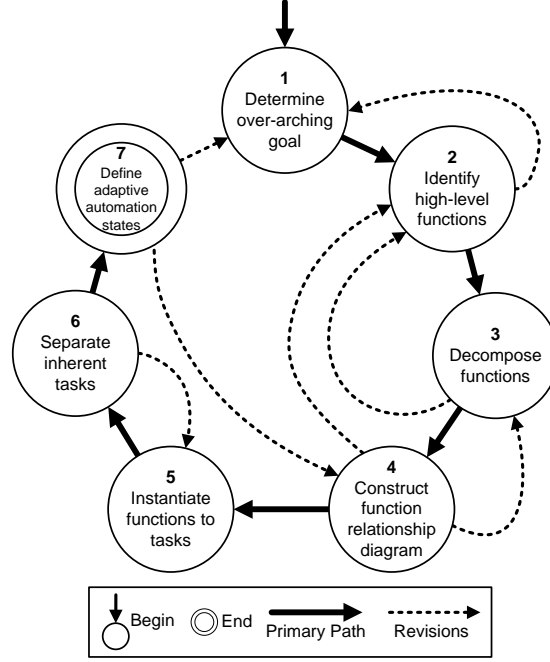
Figure 2: The Function-to-Task Design Process Model, depicting the developmental flow and typical revision loops necessary for design refinement.

3.1. Step 1: Determine Over-Arching Goal

The first step in the proposed process model, the designer determines the goal(s) of the system. The over-arching goal should answer the question, "What is the system trying to achieve?" Any predetermination as to how the task must be accomplished should be excluded. For example, a goal to "obtain milk through a purchase," contains no pre-conceived notion of *how* to purchase the milk. The overall goal should be distilled to only its essential elements–those requirements that are unavoidable. For example, obtaining milk is a less exclusive goal than purchasing milk. However, broadening the goal beyond solutions under serious consideration is counter-productive (e.g., we would not expand the goal of purchase milk unless we would consider alternate methods of obtaining milk).

3.2. Step 2: Identify High-Level Functions

The second step is to identify the functions that must be performed to achieve the goal(s). The question to answer at this stage is, "How do we achieve the over-arching goal?" The functions at this stage should be high-level, and–depending on the goal–could consist of only one function. During the successive iterations of function decomposition, these high-level functions are decomposed until they reach the atomic function level. Consequently, all functionality for which the system must account falls under a high-level function. At this point, the designer

assigns no performing entity to each function. Therefore, the high-level functions must be defined such that they can be allocated to any available entity.

3.3. Step 3: Decompose Functions

Functions are composed of sub-functions in a modular or hierarchical fashion. The complexity of a given function depends on the number and interrelationships of its sub-functions. When choosing how to automate a function, two components of the function are significant: (1) the lowest level functions (e.g., atomic functions) that make up the larger function and (2) the relationships among these atomic functions.

All non-atomic functions are composed of lower-level functions. There are many proposed methods for decomposing a function, including Integrated Computer Aided Manufacturing Definition for Function (IDEF) Modeling (Buede, 2011). The designer should perform decomposition until functions are indivisible between multiple cognitive entities, resulting in *atomic functions*. For example, if a human can perform part of a proposed "atomic function," while a second part is assigned to another human–or even a machine–that function is not yet atomic. In practice, decomposing each function to the point where it is indivisible is not necessary, but instead the designer should decompose each function to the point at which it is impractical to allocate a portion of a function to two separate entities. With system evolution, it may be necessary to readdress the function decomposition as functions which are impractical to allocate to separate entities may change as technology evolves.

The actions taken in Step 3 repeatedly address the question, "Can more than one entity perform function x?" For the purposes of system representation, step three should produce a set of nodes. Although it can be useful to use graphical depictions of the atomic functions (such as IDEF diagrams that maintain knowledge of the hierarchical decomposition (Cheng-Leong et al., 1999)) one must take care when naming the functions to insure that multiple instances of the same atomic function are named commonly without assigning multiple, different functions with a common name.

3.4. Step 4: Construct Function Relationship Diagram

Once all atomic functions are identified, the action of building the uninstantiated functions into tasks begins with exploring the relationships of the atomic functions. In order to complete a function, a subset of its atomic functions must be completed in a pre-arranged order. These relationships are depicted in a *function relationship diagram* (FRD) wherein the atomic functions represent nodes and the information transferred between nodes by the connecting edges.

Each set of atomic functions can have two possible relationships: one function relies on another (i.e. the completion or product of one task directly influences the other) or the functions are independent of each other. At this point, one might argue for the inclusion of concurrently performed tasks. However, if the set of functions remaining after decomposition are truly atomic in nature, then concurrent tasks will involve the iterative exchange of information between the

two functions. Therefore, a loop (arrows flowing in both directions) between the two functions can represent the relationship.

Multiple arrows (elements of information) may flow into or out of a given node. If an atomic function does not connect to other members of the higher-level function from which it was derived, the function decomposition should be re-addressed, as this condition violates the rules of the function decomposition. The diagram at this point should not involve the instantiation of function performers (i.e. it is still a function relationship diagram and not a task diagram).

3.5. Step 5: Instantiate functions to tasks

In step five, the system designer allocates each function to a cognitive entity: human or machine. Specific instances of humans and machines are not assigned–we are concerned only *that* a human or machine is performing the function, *not which* human or machine performs it. This step sets a baseline for the states of automation.

The first step in task instantiation involves induced assignments. Some constraint may mandate the instantiation of a specific function–or set of functions–to a specific entity. Induced assignments can come from rules, capabilities, available resources, or other avenues, but must be addressed no matter the reason for their inclusion. These are assigned first, before any other instantiations are made. Examples of induced functions include decision nodes in systems where humans hold final decision authority, or a complex calculation that a human is incapable of performing and a machine must perform.

Once the induced assignments are made, the designer can address the more flexible assignments. The adaptive automation task allocation model discussed in Section 2 enables the determination of which tasks to assign to a human or machine. By using the model demonstrated in Figure 1, tasks can be assigned to the entity capable of performing the function with maximum proficiency. Although this model may provide insight into which function nodes to instantiate to which entity, it can also draw attention to nodes that are not clearly favored to one entity or the other. Consideration of these tasks is then warranted in later stages when adaptive allocation is addressed.

The TRD should indicate each node as human or machine, with distinction for those nodes that are induced assignments. A *task relation diagram* (TRD) that demonstrates the flow of information from one entity to another results from step five.

3.6. Step 6: Separate Inherent Tasks

At this stage, all functions have been allocated to entities. The resulting TRD consists only of explicit tasks the atomic functions specify. Completion of the task allocation necessitates the specification of inherent tasks. Inherent tasks are those tasks that present themselves as the product of a specific task instantiation. However, these inherent tasks can also result from the interactions between the explicit tasks or specific resources available to the system.

The information exchange between entities during a task handoff is the primary source of inherent tasks. Once the designer assigns a task to either a

11

machine or human, a set of new complexities emerge through the task relationship diagram: task handoffs. Figure 3 demonstrates the four types of task handoffs possible: human to human, human to machine, machine to human, and machine to machine.
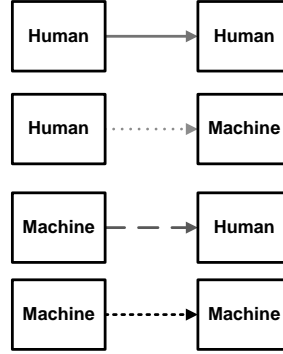

Figure 3: The four possible cognitive handoff types between two entities.

However, the most important types of task handoffs are those that cross between human and machine. A human-to-machine or machine-to-human task handoff requires two inherent tasks that are not present in the underlying functions: (1) formatting and communicating the information by the *losing entity* and (2) perception of the information by the *gaining entity*.

Communication of information requires the current performing entity to format the information such that the next entity understands it. For example, a machine that just completed a movie recommendation search task must ensure that it communicates the recommended films to the human before the human can complete the subsequent movie selection task (i.e., displays this information on a screen). On the other end, perception involves the next task performer's ability to obtain and interpret the information communicated to permit subsequent task completion. It's important to make the inherent task nodes in the task relationship diagram visually distinctive. This distinction permits complex task relationships to become more apparent. Step six produces a complete TRD similar to that produced by the previous step, but including both explicit and inherent tasks.

At this point, the designer may find it useful to reiterate through the process to ensure that the diagram truly represents the desired process and system. After this stage, an initial allocation exists. Modeling or prototyping tools can then be used to determine if the human or humans assigned to operate within the system are capable of performing the tasks required from them during typical system operation while having high enough workload to remain engaged with the system. If not, steps five and six are revised until the design attains a desired level of workload. To reduce workload, for example, the human can give a task involved within a complex relationship within the TRD to the machine.

12

3.7. Step 7: Define Adaptive Automation States

The inherent tasks of communication and perception, provide one of the most important steps in designing an automation system, due to the complexity they add to a system. When the designer adds adaptive automation to the system, an understanding of cognitive task handoffs is crucial. The selection of a set of atomic tasks–or groups of tasks–in the TRD to become adaptive nodes makes the automation adaptive, by identifying nodes that can switch between human and machine instantiation based on some pre-defined trigger. The TRD aids in identifying potentially useful locations for adaptive nodes.

Selecting a given node as an adaptive node, based on the TRD, adds complexity to the overall system. Figure 4 shows an updated version of Figure 3, wherein all possible relationships are represented when at least one of the nodes is an adaptive node.
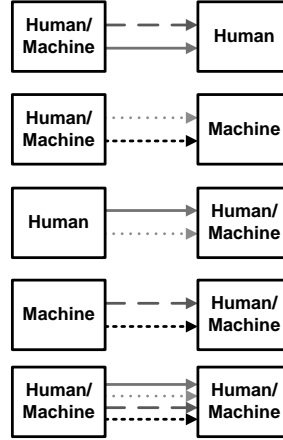

Figure 4: The four possible cognitive handoff types between two entities.

As a node switches from static human or machine to adaptive, the number of handoff types needed approximately doubles for each outgoing connection. Figure 5 illustrates this principle, where the number of handoff types present in the entire diagram goes from four in the original to seven in the resultant diagram. Furthermore, changing only one node from 'Machine' to 'Human or Machine' achieves this increase.

Step seven finalizes the definition of adaptive nodes, and the TRD aids analysis. Five analysis tools include: determining the number of possible states, node clustering, task handoff analysis, branch counting, and inherent task load comparison. By iterating through these tools, an adaptive automation system emerges.

3.7.1. Number of possible states

Once the designer selects adaptive nodes, they must readdress the complexity and handoffs created through the selection. One way to look at complexity

13

involves determining the number of possible automation states. For each adaptive automation node in the relationship graph, two possible states exist: human and machine. Therefore, the number of possible states equals 2^x, where x is the number of adaptive nodes in the current design. For example, the model on the left of Figure 5 contains one possible state, while the model on the right contains two.

3.7.2. Node clustering

Functions clustered based upon the degree of the edges in and out of a given node tend to provide similar or highly inter-related functions. Therefore, automation of the cluster as a group can often be achieved with greater effect than just automating one function in the group. Conversely, the designer could also instantiate all of the tasks in a cluster to a human, since the cognition will not need to change. An example of this would be in piloting an aircraft. Although takeoff function contains many lower level functions, there are many complex groups of functions within it that naturally group together to ensure a proper amount of situational awareness–provided through the right kinds of feedback.

3.7.3. Task handoffs

One way the TRD can help analyze the designed system is through an analysis of the task handoffs, specifically the number of different-entity handoffs. In each possible case where a machine hands off to a human or human to a machine, count one handoff. In the case where a node is set as adaptive, this implies that a handoff from an adaptive node to another adaptive node counts twice, while an adaptive to non-adaptive node handoff will counts once. This handoff count suggests the number of nodes where cognitive load shifts from one entity to another. By focusing on these nodes, potential bottlenecks appear due to certain handoff tasks taking place more often than at other locations. Highlighting all of these tradeoffs allows the system designer to visualize the
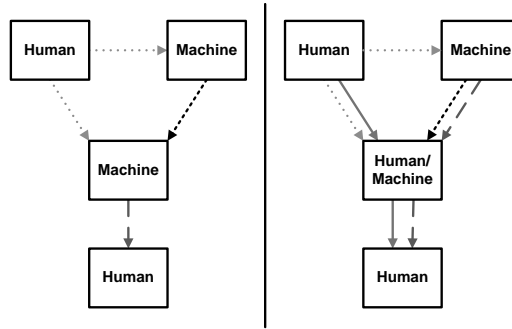

Figure 5: Complexity added to automation design considerations by changing the cognitive function operator from a hard assignment (machine) to an adaptive assignment (human/machine). The number of functional handoff types doubles for each handoff involved.

locations where inherent tasks–specifically those associated with communication and perception, as described in Section 3.6–reside.

3.7.4. Branch counting

Branch counting refers to the idea of determining the number of other atomic tasks to which one specific atomic task connects. A task that influences or is influenced by a large number of tasks makes automation more difficult. Tasks that have large branch counts can often make good candidates for node clustering. Conversely, single branches within a TRD can often indicate good places to put adaptive nodes, as the inherent task load will (likely) be smaller.

3.7.5. Inherent task load comparison

An inherent task load comparison provides another way to analyze the effectiveness of different designs. This consists of a comparison of the relationship diagrams created when the TRD instantiates one function as a human task versus when the TRD instantiates the same function as a machine task. The main difference that will most likely arise in this type of comparison is the number of tasks that are added or subtracted due to a specific instantiation. A comparison of the two instantiations, helps to visually communicate inherently understandable design decisions.

4. Function to Task Process Illustrated

The design of an automated route creation game–titled Space Navigator, shown in Figure 6–inspired by others in the genre such as Harbor Master (Imangi Studios, L.L.C., 2011), Flight Control (Firemint Party, Ltd., 2011), and Martian Control (Beebe and Beebe, 2010) illustrates the proposed function-to-task process model. The game involves spaceships appearing on a computer screen that a player must select and drag to a specific destination planet, while avoiding obstacles and acquiring point bonuses. The game consists of a ten-minute session, during which the goal is to score as many points as possible. The player can earn positive points in two ways: landing a spaceship on its destination planet and "picking up" a randomly appearing alien (i.e. the ships path traverses over the alien's location on the screen). The player loses points in two ways: permitting two ships collide (the ships are also destroyed) and permitting a spaceship to enter a no-fly zone. No-fly zones are rectangles that appear and disappear at random intervals and points are lost for every second the ship remains in the zone. Ships appear at random intervals and can come into the screen from any direction.

Within this game, the player's workload varies with changes in the rate of appearance of spaceships, aliens and no-fly zones, as well as the number of planets. The system employs adaptive automation to permit the player to effectively increase points when the number of active entities–planets, spaceships, aliens, no-fly zones–increases beyond a rate that the player can control. The time and location that entities appear are selected randomly based upon probability distributions and are thus not available to the automation system. The automation
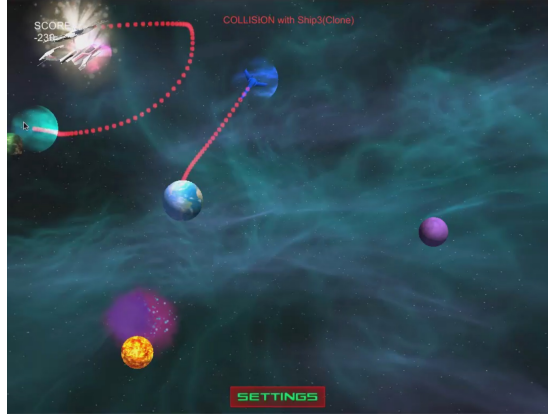
Figure 6: Screenshot from an instance of the Space Navigator game.

system can aid the player based upon temporally-available information but has no knowledge of future events. Further, the system has limited processing power and is capable of determining routing information for entities, but determines the route for a single entity at a time. Once the performing entity selects a route, the automation might identify an impending collision and recommend a change in route information. However, this example assumes that the human player must permit a change in any route once an initial route is established.

4.1. Goals and Functions

To fulfill Step 1, we ask the question "What are we trying to achieve?" For this game, the goal is to score the most possible points in the allotted time. The high-level functions of the game then represent the way that points are scored. Four high-level functions become apparent:

1. Land a spaceship on its destination planet.
2. Pick up an alien.
3. Avoid collisions with other spaceships.
4. Avoid traversing no-fly zones.

Answering the question, "Can we further divide function x?" can help decompose these four functions. After asking the question we obtain the following list of atomic functions:

1. Function 1: Move spaceship to intended target planet.

   - Determine the best ship to draw route.
   - Identify destination planet of selected ship.
   - Identify if ships have routes already.
   - Create a set of possible routes.
   - Select a route.

16

- Draw a line from selected ship to destination.

2. Function 2: Pick up aliens

- Identify all available, non-selected aliens.
- Identify destination planet of selected ship.
- Determine if route change to pick up alien is worth points gained.
- Determine if selected ship has a route already.
- Adjust route to pick up alien.

3. Function 3: Avoid other spaceships

- Detect likely collisions.
- Identify destination planet of selected ship.
- Determine if selected ship has a route already.
- Determine if route change to avoid collision is worth points gained
- Adjust route to avoid collisions.

4. Function 4: Avoid no-fly zones

- Identify no-fly zones.
- Identify ships headed toward a no-fly zone.
- Identify destination planet of selected ship.
- Determine if selected ship has a route already.
- Determine if no-fly zone traversal is worth lost points.
- Adjust route around no-fly zone.

This atomic function list demonstrates two important concepts previously discussed in Section 3.3: the circumstance-specific nature of atomic functions and the overlap of specific atomic functions. For the Space Navigator game, the atomic functions listed above could be considered more complex functions depending on your interpretation of the process. For example, under the avoid no-fly zones high-level function, the atomic function "Determine if no-fly zone traversal is worth lost points" could be considered a non-atomic function made up of sub-functions such as determine the number of potential points lost, determine amount of time added, determine increased collision likelihood, etc. The FRD retains this function as an atomic function as a human could somewhat intuitively weight the costs of traversing a no-fly zone, taking into account extenuating circumstances. The second concept is that atomic functions appear in multiple locations within the hierarchy. For example, the atomic function "identify destination planet of selected ship" appears in all high-level functions and is the same function in all cases.
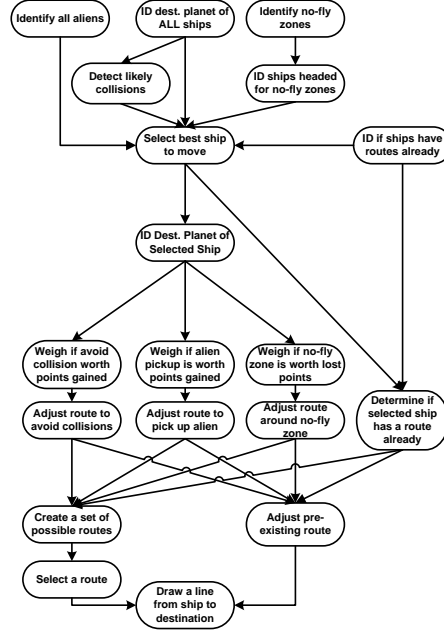
Figure 7: Functional relationship diagram of the Space Navigator game.

4.2. Construct Functional Relationship Diagram

We now produce the functional relationship diagram. We analyze each unique atomic function in relation to all of the other functions and assigned relationships (independent or reliant) based upon the transfer of information from one function to another. The end result of this relationship mapping is the function relationship diagram shown in Figure 7. Two concepts are illustrated well in this diagram.

First, there are a few instances where multiple higher-level functions contain the same atomic function (e.g. "identify destination planet of selected ship"). Only one node in the functional relationship diagram represents these functions. However, these functions interact with many different functions. The "identify destination planet of selected ship" function directly influences three separate functions. Therefore, functions that overlap multiple higher-level functions provide potential bottlenecks in the relationship diagram. That is, the information these functions produce must be available to any human or machine entity to permit subsequent functions' performance.

Secondly, relationships that seem compartmentalized in the function decomposition can appear highly interconnected in the relationship diagram. The four high-level functions identified for the Space Navigator game are separated distinctly in the functional decomposition in Section 4.1. However, when they are placed into relationship with each other, the sub-functions provide a system that cannot be easily divided along the lines of the previously defined high-level

functions. This change in perspective can also influence a different understanding of the high-level functions themselves. It may be the case that the high-level functions can be defined differently based on their structure in the functional relationship diagram. This demonstrates why we must continually refine the design of the system and readdress previously completed steps as needed.

4.3. Determine Function Allocation

The design goal in this example is to apply automation to aid the user when interacting with this game where the assumed default state is that the human operator will perform all functions. Therefore, the goal of the function allocation in this particular example is to identify alternate automation states. Referring to Figure 7, one can see that the functions in the center of the diagram are highly inter-connected. This interconnection implies that the human and machine would need to exchange significant amounts of information if elements within this region of the figure were divided between these entities. However, other elements near the periphery of the diagram are not as highly interconnected. As a result, allocation of many of these elements to the machine are likely to result in less need for communication between the human and machine.

Based on this analysis and the performance of the human and machine, Figure 8 represents a potential task allocation and resulting task relationship diagram. In the diagram, tasks that the human controls are shown in blue and those in red are alternatively allocated to the machine. As discussed earlier, the "C/P" nodes are also shown, indicating the need for the entity performing the function near the "C" node to perform the inherent task of formatting information for display to the receiving entity and the need for the entity performing the function near the "P" node to perform the inherent task of perceiving and interpreting the information to enable performance of the function.

4.4. Show Inherent Tasks

Because of the selected function allocation, there are several task handoffs from human to machine and vice versa as the C/P nodes indicate. Some of the communication/perception chains are inconsequential, like communicating from the machine to human if a specific spaceship has a route already–as a simple path is already drawn between entities to aid transfer of this information. However, others are more difficult. For example, communicating the destination planet for all ships to a human can be simple, but it is important and perhaps difficult to ensure perception. If the ships are color-coded to align with a specific planet, this task is simple for most people when few entities are available. However, it becomes increasingly difficult as the number of entities increase and in some cases impossible for certain individuals (e.g., those who are color blind). Therefore, it is not only needed to communicate the information, but to confirm the transfer of critical information to insure a handoff.
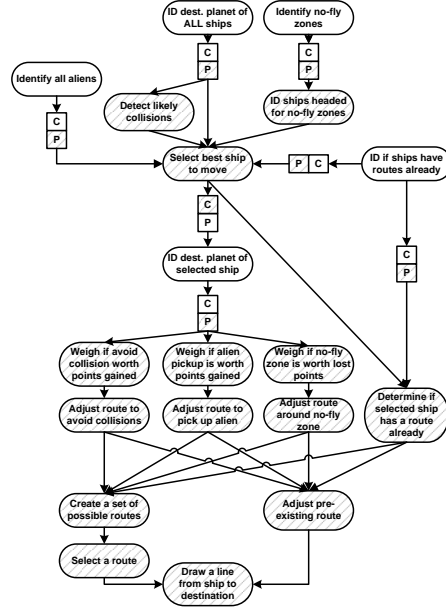
Figure 8: Task relationship diagram of the Space Navigator game, where blue nodes have been allocated to the human and red to the machine. The C/P nodes represent nodes where a task handoff from machine to human or vice versa and a communication-perception task must take place.

4.5. Create Adaptive Automation

Finally, we apply adaptive automation to the TRD. Figure 9 shows how the process changes after adaptive automation nodes are selected. The most important areas to address in the adaptive automation process for the Space Navigator game were the node clustering, branch counting, and inherent task load counting. Node clustering in this example was used to convert a large block of tasks as one larger task. The purple box surrounding the adaptive atomic tasks represents a clustering of the tasks into a larger "ship selection" task. This node cluster was not evident without the TRD, as the functional decomposition was performed based on ways that reach the goal (i.e. score points). The node clustering was aided first by counting branches. It was obvious that the "Select best ship to move" node as well as all the "weigh if..." nodes contained high branch counts. However, the boundary around the cluster was placed such that each crossing contained only one crossed branch. The "ID Destination planet of selected ship" node only has one exiting branch, as it will communicate the information to the human for the purpose of all ensuing tasks in the same way.

An inherent task comparison shows that there was a slight change based on the inherent tasks added in comparison to the TRD in Figure 8. The "Select best ship to move" node now has a "C/P" node coming out of it. However, by clustering the nodes as selected in Figure 9, when all nodes are set to machine,
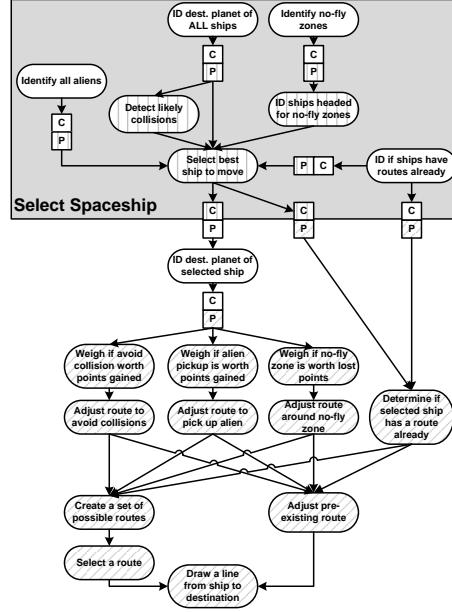
20

Figure 9: Task relationship diagram of the Space Navigator game, where blue nodes have been allocated to the human, red to the machine, purple as adaptive automation nodes. The C/P nodes represent nodes where a task hands off from machine to human or vice versa and a communication-perception task must take place.

there will be fewer inherent tasks added than with the original set of nodes.

5. Conclusions

The presented function-to-task design process model creates a set of visual diagrams enabling designers to better allocate tasks between human and machine. This is achieved through a set of five analysis tools allowing designers to identify points within a function network where the transitions between human and machine entities can facilitate adaptive automation. This paper proceeds as follows. Section 2 reviews the design processes currently in place for adaptive automation systems. Section 3 presents the function-to-task design process model. Section 4 illustrates the function-to-task design process model through a system design iteration.

This paper presents an investigation of the effects of explicit and inherent task load to provide a design process model for adaptive automation systems. The resulting function-to-task design process model demonstrates that adaptive automation requires the dynamic allocation of discrete functions to the human or machine rather than adjusting the degree of automation on a continuum. Further, the process model demonstrates that reallocation of functions imposes a change in implied tasks to permit the proper exchange of information between

the performing entities. As such, reallocation of a function implies a change in information flow between the machine and human and this change requires the user to perform tasks which require cognitive and physical resources to obtain and communicate this information. The function-to-task design process model prescribes the steps necessary to create the function relationship diagram and task relationship diagram, enabling designers to better allocate tasks between human and machine. These allocation improvements are achieved through a set of five analysis tools that allow designers to identify points within a function network where the transitions between human and machine entities enable adaptive automation.

It has been previously suggested that each of these elements of information are necessary to support allocation (Wright et al., 2000), although they are often represented in separate diagrams. Within adaptive automation system design, this representation is particularly important as it permits the depiction of not only explicit but inherent tasks necessary for the conveyance of information between cognitive entities. Consideration of the information available in the task relationship diagrams when performing task allocation permits the designer to understand and potentially reduce the volume or complexity of information exchange between a human and machine. This tool may also help to reduce unwanted redundancy between the functions the human and the machine perform by clarifying the form of the information necessary to facilitate human decision making.

Acknowledgments

References

Beebe, J., Beebe, B., December 2010. Flight Control clone with Corona in 30 hours.
URL `http://jonbeebe.tumblr.com/post/2301674980/flight-control-clone-with-corona-in-30-hours`

Buede, D. M., 2011. The engineering design of systems: models and methods. Vol. 55. John Wiley & Sons.

Bye, A., Hollnagel, E., Brendeford, T. S., 1999. Human–machine function allocation: a functional modelling approach. Reliability Engineering & System Safety 64 (2), 291–300.

Byrne, E. A., Parasuraman, R., 1996. Psychophysiology and adaptive automation. Biological Psychology 42 (3), 249–268.

Calleja, J. A. B., Troost, J., 2005. A fuzzy expert system for task distribution in teams under unbalanced workload conditions. In: Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on. Vol. 1. IEEE, pp. 549–556.

Cheng-Leong, A., Li Pheng, K., Keng Leng, G. R., 1999. Idef*: a comprehensive modelling methodology for the development of manufacturing enterprise systems. International Journal of Production Research 37 (17), 3839–3858.

Colombi, J., Miller, M., Schneider, M., McGrogan, J., Long, D., Plaga, J., 2012. Predictive mental workload modeling for semi-autonomous system design: Implications for systems of systems. Journal of Systems Engineering 15 (4), 448–460.

Concepts, I., Group, T. W., 1998. Incose se terms glossary. Tech. rep., International Council on Systems Engineering, Seattle, WA.

Dorneich, M. C., Ververs, P. M., Mathan, S., Whitlow, S., Hayes, C. C., 2012. Considering etiquette in the design of an adaptive system. Journal of Cognitive Engineering and Decision Making 6 (2), 243–265.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., Beck, H. P., 2003. The role of trust in automation reliance. International Journal of Human-Computer Studies 58 (6), 697–718.

Endsley, M. R., 1999. Level of automation effects on performance, situation awareness and workload in a dynamic control task. Ergonomics 42 (3), 462–492.

Feigh, K. M., Dorneich, M. C., Hayes, C. C., 2012. Toward a characterization of adaptive systems: A framework for researchers and system designers. Human Factors: The Journal of the Human Factors and Ergonomics Society.

Firemint Party, Ltd., December 2011. Flight Control. https://itunes.apple.com/us/app/flight-control/id306220440?mt=8.

Fitts, P., Chapanis, A., Frick, F., Garner, W., Gebhard, J., Grether, W., Henneman, R., Kappauf, W., Newman, E., Williams, Jr., A., 1951. Human engineering for an effective air-navigation and traffic-control system. Tech. rep., National Research Council Committe on Aviation Psychology, Washington, D.C.

Geiselman, E. E., Johnson, C. M., Buck, D. R., Patrick, T., 2013. Flightdeck automation: A call for context-aware logic to improve safety. Tech. rep., Air Force Institute of Technology.

Haarmann, A., Boucsein, W., Schaefer, F., 2009. Combining electrodermal responses and cardiovascular measures for probing adaptive automation during simulated flight. Applied Ergonomics 40 (6), 1026–1040.

Imangi Studios, L.L.C., April 2011. Harbor Master. `https://itunes.apple.com/us/app/harbor-master/id313014213?mt=8`.

Itoh, M., 2011. A model of trust in automation: Why humans over-trust. In: SICE Annual Conference (SICE), 2011 Proceedings of. IEEE, pp. 198–201.

Itoh, M., Abe, G., Tanaka, K., 1999. Trust in and use of automation: their dependence on occurrence patterns of malfunctions. In: Systems, Man, and Cybernetics, 1999. IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on. Vol. 3. IEEE, pp. 715–720.

Kaber, D. B., Riley, J. M., Tan, K.-W., Endsley, M. R., 2001. On the design of adaptive automation for complex systems. International Journal of Cognitive Ergonomics 5 (1), 37–57.

Kidwell, B., Calhoun, G. L., Ruff, H. A., Parasuraman, R., 2012. Adaptable and adaptive automation for supervisory control of multiple autonomous vehicles. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Vol. 56. SAGE Publications, pp. 428–432.

Kirwan, B., 2005. Developing human informed automation in air traffic management. Contemporary Issues In Human Factors And Aviation Safety, 247.

Lee, J. D., See, K. A., 2004. Trust in automation: Designing for appropriate reliance. Human Factors: The Journal of the Human Factors and Ergonomics Society 46 (1), 50–80.

MacDonald, W., 2003. The impact of job demands and workload on stress and fatigue. Australian Psychologist 38 (2), 102–117.

Manzey, D., Reichenbach, J., Onnasch, L., 2012. Human performance consequences of automated decision aids the impact of degree of automation and system experience. Journal of Cognitive Engineering and Decision Making 6 (1), 57–87.

Merriam-Webster, 2013. Adapt.
URL `http://www.merriam-webster.com/dictionary/adapt`

Merritt, S. M., Heimbaugh, H., LaChapell, J., Lee, D., 2012. I trust it, but i dont know why effects of implicit attitudes toward automation on trust in an automated system. Human Factors: The Journal of the Human Factors and Ergonomics Society.

Nakazawa, H., 1993. Alternative human role in manufacturing. AI & society 7 (2), 151–156.

Parasuraman, R., Cosenzo, K. A., De Visser, E., 2009. Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. Military Psychology 21 (2), 270–297.

Parasuraman, R., Sheridan, T. B., Wickens, C. D., 2000. A model for types and levels of human interaction with automation. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 30 (3), 286–297.

Parasuraman, R., Sheridan, T. B., Wickens, C. D., 2008. Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. Journal of Cognitive Engineering and Decision Making 2 (2), 140–160.

Parasuraman, R., Wickens, C. D., 2008. Humans: Still vital after all these years of automation. Human Factors: The Journal of the Human Factors and Ergonomics Society 50 (3), 511–520.

Prevot, T., Homola, J., Mercer, J., 2008. Human-in-the-loop evaluation of ground-based automated separation assurance for nextgen. In: Congress of International Council of the Aeronautical Sciences Anchorage, Anchorage, AK.

Price, H. E., 1985. The allocation of functions in systems. Human Factors: The Journal of the Human Factors and Ergonomics Society 27 (1), 33–45.

Reid, G. B., Nygren, T., 1988. The subjective workload assessment technique: A scaling procedure for measuring mental workload. Human mental workload 185, 218.

Ross, D. T., 1977. Structured analysis (sa): A language for communicating ideas. Software Engineering, IEEE Transactions on (1), 16–34.

Rouse, W. B., 1977. Human-computer interaction in multitask situations. Systems, Man and Cybernetics, IEEE Transactions on 7 (5), 384–392.

Rouse, W. B., 1981. Human-computer interaction in the control of dynamic systems. ACM Computing Surveys (CSUR) 13 (1), 71–99.

Sheridan, T. B., Parasuraman, R., 2005. Human-automation interaction. Reviews of human factors and ergonomics 1 (1), 89–129.

Sheridan, T. B., Verplank, W. L., 1978. Human and computer control of undersea teleoperators. Tech. rep., DTIC Document.

Valente, A., Carpanzano, E., 2011. Development of multi-level adaptive control and scheduling solutions for shop-floor automation in reconfigurable manufacturing systems. CIRP Annals-Manufacturing Technology 60 (1), 449–452.

Valente, A., Carpanzano, E., Nassehi, A., Newman, S. T., 2010. A step compliant knowledge based schema to support shop-floor adaptive automation in dynamic manufacturing environments. CIRP Annals-Manufacturing Technology 59 (1), 441–444.

Wickens, C. D., 2008. Multiple resources and mental workload. Human Factors: The Journal of the Human Factors and Ergonomics Society 50 (3), 449–455.

Woods, D. D., Cook, R. I., 2006. Incidents–markers of resilience or brittleness. Resilience Engineering. Concepts and Precepts. Ashgate, Aldershot, UK.

Wright, P., Dearden, A., Fields, B., 2000. Function allocation: a perspective from studies of work practice. International Journal of Human-Computer Studies 52 (2), 335 – 355.
URL `http://www.sciencedirect.com/science/article/pii/S1071581999902926`

Vitae

Jason Bindewald received his B.A. degree in Computer Science from Gettysburg College, Gettysburg, Pennsylvania, in 2005. He received his M.S. degree in Cyber Operations in 2011 and is currently a Ph.D. Student in the Department of Computer and Electrical Engineering at the the Air Force Institute of Technology. He is a member of the Tau Beta Pi engineering honor society. His research interests are in the areas of adaptive auotmation, artificial intelligence, games, and machine learning with emphasis on human-machine interaction.

Michael E. Miller (PhD Virginia Tech) is an Assistant Professor at the Air Force Institute of Technology. His research interests include human systems integration, modeling and measurement of human performance and human interface design. Prior to joining AFIT, Dr. Miller spent more than 15 years with Eastman Kodak Company as a systems and human factors engineer. Dr. Miller holds over 80 US Patents in digital imaging and OLED systems. He is a senior member of SID, a member of the Human Factors and Ergonomics Society, a senior member of the IIE, and a member of INCOSE.

Gilbert L. Peterson is an Associate Professor of Computer Science at the Air Force Institute of Technology, and Vice-Chair of the IFIP Working Group 11.9 Digital Forensics. Dr. Peterson received a BS degree in Architecture, and an M.S and Ph.D in Computer Science at the University of Texas at Arlington. He teaches and conducts research in autonomous systems, digital forensics, and statistical machine learning. His research has been sponsored by the NSF, DARPA, AFOSR, AFRL, and JIEDDO. He has over 75 peer reviewed publication, and a book. In 2008, he received the Air Force Junior Scientist of the Year Category I award.

Appendix D

Published Article: Splawn, J. and Miller M.E. (2013) Prediction of perceived workload from task performance and heart rate measures, Proceedings of the Human Factors and Ergonomics Society, San Diego, CA.

PREDICTION OF PERCEIVED WORKLOAD FROM TASK PERFORMANCE AND HEART RATE MEASURES

Joshua M. Splawn and Michael E. Miller, PhD
Air Force Institute of Technology

To improve operator efficiency and effectiveness, designers increasingly apply automation to allocate tasks once performed by human operators to the system. Unfortunately, these systems are often complex, potentially imposing increased mental task load on the operator, or placing the operator in a supervisory role where they can become overly dependent on automation. A proposed solution is adaptive automation, which increases automation when an operator is overloaded, and disabled as the operator has spare mental capacity. Changes in performance and physiological measures have shown promise in triggering changes in automation levels. However, the literature lacks well-documented or consistently supported measures for mental workload prediction. The present work sought to define a model which could predict perceived workload as a function of performance and heart rate measures by imposing various levels of task loading on a group of individuals while monitoring their performance, recording their heart rate information with an electrocardiogram and obtaining subjective estimates of mental workload. Heart rate (HR) and several heart rate variability (HRV) measurements where significantly affected by Task Load. This paper describes a linear regression model for predicting participants' perceived workload as a function of a proposed summary performance metric and HR measures.

INTRODUCTION

Present Air Force (AF) systems are generally more complex and require the operator to perform more tasks than previous systems. This description is true for Remotely Piloted Aircraft (RPA. These aircraft are now one of the most demanded capabilities that the USAF presents to the Joint Force (USAF, 2009), having undergone explosive growth despite their complexity. Specifically the complexity of battlefield operations is increasing as new technologies to address changing threats. Often, operators can still function at an optimum level; however, if the task load becomes too great operator performance can suffer (de Waard, 1996). Consequently, technologies and techniques are needed to assist RPA operators in maintaining optimal performance.

One possible method of aiding operators is through adaptive automation (AA). AA is "a system of controlling flexibly and dynamically the allocation of tasks between human operators and computer systems in complex multi-task environments" (Tattersall & Fairclough, 2003). Adaptive automation abandons the fixed allocation of tasks between machine and operator in favor of adjusting this allocation during system operation (Parasuraman & Hancock, 2008). These systems seek to balance mental workload, maintain maximum operator situational awareness, permit the operator to maintain cognitive skills, and help the operator to gain trust in the automation (Parasuraman and Byrne, 2003). Parasuraman discusses five methods to trigger this automation, which include: critical events, operator performance, operator physiological assessment, operator modeling, and a hybrid of these techniques (Parasuraman, 2003). The two techniques that have shown promise in triggering changes in adaptive automation within several studies are operator performance and physiological measures (Parasuraman, 2003). However, research which has explored

these measures has not provided robust, repeatable models of human mental workload as a function of these variables.

Throughout the years, human factors engineers have employed many techniques to quantify and measure workload. These techniques vary greatly, from purely subjective questionnaires (of which there are many), that require the participant to measure their own perceived workload, to objective physiological measures such as cardiovascular responses, oculometry, galvanic skin response, and fMRI (Booher, 2003). In particular, cardiovascular measurements (HR and HRV) have shown correlation with mental workload levels in many studies (Averty, Athenes, Collet, & Dittmar, 2002; de Waard, 1996; Wickens & Hollands, 2000). Unfortunately, combined models, which permit prediction based upon performance and heart rate measures, are not readily available. As a result, the present study develops a model to predict workload as a function of task performance and heart rate measures. This model should provide a trigger for automation in future AA studies.

BACKGROUND

To implement adaptive automation to balance mental workload, it is important to understand mental workload and methods to measure it. There are several definitions of workload in the literature, but the underlying theme is that mental workload is the amount of mental effort needed to accomplish a task or goal. One definition determines workload to be "the amount of cognitive or attentional resources being expended at a given point in time" (Booher, 2003), while another states, "Workload is a general term used to describe the cost of accomplishing task requirements for the human element of man-machine systems" (Tsang & Vidulich, 2003). These definitions each focus on the resources expended by the operator of the system at a given time or under a time constraint. Inherent in these definitions is the

need to measure mental effort on an ongoing basis throughout the operator's shift. While one might directly observe physical workload, the measurement of mental workload often requires indirect means. These indirect measures include performance on primary or secondary tasks, subjective measures, or electrophysiological measures (Booher, 2003).

Primary task performance measures involve assigning an operator one or more tasks, placing emphasis on the speed and accuracy of the most important task. While these measures are useful and can be collected continuously to assess workload level, most real world environments require the operator to perform multiple, often disjoint tasks, which must be summarized to determine overall task performance.

Parasuraman and colleagues applied primary task performance to trigger changes in level of automation within adaptive automation systems (Parasuraman, Cosenzo, & De Visser, 2009). This study required the operator to interact with an unmanned, ground-vehicle, control station to perform simultaneous tasks. These tasks included completing specific actions at various waypoints depending on the situation, as well as acknowledging their call sign if presented, and responding to a situation awareness probe. Finally, the operators were asked to respond to a change detection task and hit the space bar when they noticed a change. The operator's performance for this final task was used to determine whether or not automation should be invoked. The results showed that the adaptive automation increased performance on only the final task, but it reduced the participants' subjective assessment of workload.

Appropriate physiological measures are also appealing as a potential trigger as these measures do not negatively influence performance (Wickens & Hollands, 2000). These measures provide continuously gathered data, and the measures often correlate with workload. The literature demonstrates that many physiological can provide a reliable indication of workload. However, this research will primarily focus on cardiovascular measures. Heart-rate variability (HRV) is a specific cardiovascular measure that often correlates with workload. It has been stated that; "In general HRV decrease is more sensitive to increases in workload than heart rate (HR) increase" (de Waard, 1996). Many studies have found that HRV is sensitive to a number of different difficulty manipulations (de Waard, 1996).

HRV has also been applied to trigger automation. Parasuraman conducted a study in which two groups of participants used a modified version of MAT-B, called EICAS-MAT, for 90-minute sessions. The session was divided into three consecutive phases of high, low, and high difficulty with adaptive automation being provided for the high difficulty phases for one group; however, this automation was only triggered when heart rate variability (HRV) was reduced below a specific point (Parasuraman, 2003). The second group did not receive any aid from the automation, regardless of their heart rate measures. The results showed that the adaptive group had higher HRV, which is indicative of lower workload, and that their tracking performance was superior to the control group (Parasuraman, 2003). While this study did not discuss the operators' performance on the remaining three three tasks within this simulated environment,

one might assume that their was no difference between groups in those tasks and that adaptive automation employing HRV was successful in reducing workload and improving task performance.

Although HRV shows promise as a predictor of workload for adaptive automation, it is not the only potential cardiac measure. Further, HRV and related measures have limitations. Unfortunately, HRV responds to both mental workload and physical workload (de Waard, 1996; Novak, Mihelj, & Munih, 2010). Additionally, HRV is difficult to compute in real-time and can be sensitive to noise. Another limitation of all cardiac measures is the operator's acceptance, as ECG connections can be uncomfortable for day-to-day missions. However, several techniques have been discussed to remotely measure heartbeat using imaging devices (Kamshilin, Miridonov, Teplov, Saarenheimo, & Nippolainen, 2011), which may eliminate this concern in the long term.

Other cardiovascular measures include SDNN, which is the standard deviation of the inter-beat intervals (IBIs), CVRR, which is the SDNN divided by the average IBI value. Both of these measures have shown correlation with task loads (Kawakita, Itoh, & Oguri, 2010). Some of these measures recognize that heart rate can vary significantly between individuals and, therefore, require that a baseline, e.g., the research must establish a measure of resting heart rate.

One might consider combining primary task performance measures and physiological assessment to form an integrated method for mental workload assessment since each measure responds to mental workload.

METHODS

Participants

Thirteen participants participated in this study, including 4 females and 9 males. The average age was 25 years, and all were in good health. The participants included volunteer military and government civilian personnel. Approximately half of the participants completed one of two medium workload conditions. Half of each of these subgroups experienced a different baseline condition.

Equipment

Participants interacted with the Air Force Multi-Attribute Test Battery (AF_MATB), running on a laptop computer. The AF_MATB, provided a method to manipulate an operator's task load and impose different levels (high, med, low) mental workload (Miller, 2010). The original MATB software has become a mainstay for psychological research regarding cognitive workload and the most recent version (Version 2.4), which was employed in this study, has been updated to be compatible with modern operating systems (Miller, 2010). Participants used a standard laptop keyboard in addition to a USB joystick to perform the given tasks.

The AF_MATB requires the operator to perform four simultaneous tasks that simulate tasks analogous to those a flight crewmember would encounter. Tasks include system monitoring, tracking, communication, and resource allocation.

System monitoring consists of monitoring four gauges and two lights, to which the participant provides corrective action via the keyboard. A joystick controlled the cursor in the tracking task. The objective of the tracking task is to keep the unstable crosshairs within a designated rectangular target area. Communication requires the participant to listen for the appropriate call sign and change the frequency for one of four channels via the keyboard. For the resource allocation task, participants are responsible for turning on/off eight pumps to maintain a desired level (2500 +/- 300 for this research) in two main tanks in a constantly changing environment.

Integrated within AF_MATB is a subjective workload assessment scale, specifically the NASA Task Load Index (TLX). This subjective workload scale has been used in studies as an effective means to measure subjective impression of workload (Stanton, 2005). The six components of the NASA TLX scale, which are used to form an aggregate measurement of workload, are mental demand, physical demand, temporal demand, performance, effort, and frustration. A BIOPAC 150 with ECG 100C amplifier measured the electrical signals associated with the beat of the human heart.

Procedure

Participants read and signed the informed consent document. The participants attached the ECG leads to their chest under supervision of a trained experimenter having the same gender as the participant and reapplied any clothing. The participant then sat in front of the workstation and the experiment attached the ECG leads to the amplifier. The first group of participants completed the baseline by simply relaxing and performing a set of calming mental exercises. However, the baseline recordings were not reliable and often resulted in HRV values that were smaller than HRV values obtained for the low workload conditions. As a result, the second group of participants viewed a series of natural images during baseline recording. These images were selected to induce relaxation for a broad range of participants (Fedorovskaya et al., 2001).

The AF_MATB test scenarios began after completion of the baseline measurements, beginning with 15 minutes of training, during which the participant interacted with the system and received positive guidance from the experimenter. After this training, participants self determined whether to continue training or proceed with the experiment. Once the participants were comfortable with the software, they completed three, five-minute AF_MATB sessions at high, medium, and low difficulty levels. By randomizing the difficulty levels, the experimenters attempted to minimize learning effects. When each session ended, the participant completed the NASA TLX subjective workload questionnaire. The participant completed this sequence a second time. The participants were debriefed and the experiment terminated.

Data Collection and Analysis Preview

The AF_MATB software recorded and output all performance data. This data includes the response times to errors for the system-monitoring task and the communications task, but shows the root mean square (RMS) time for the other two tasks. However, as it was the goal of the experiment to correlate overall performance with the subjective workload values, the participant's performance on the four individual task scores need to be combined into a single score for each task load level. Calculation of the ratio of the time each measure was in the correct state to the time it could have been in the correct state served as the performance score (Splawn, 2013). Note that this process collapses performance on all tasks to a single score, rather than relying only on the performance of a single task as is common in the literature.

AF_MATB collects the subjective workload data input by the participants and converts it to the NASA TLX weighted workload level (WWL), which is the participant's composite score for perceived mental workload. An ANOVA was performed on the WWL to determine if it is an effective method to accurately measure the operator's perceived mental workload; i.e., if it was sensitive to the changes in task load that were imposed during the experiment.

Analysis of the cardiac measures involved AcqKnowledge version 4.0 from BIOPAC . This software allows the data to be filtered, automatically locates the ECG waveforms, identifies their peak R waves, and then runs analysis to compute heart rate and the different spectral components of heart rate variability (HRV), including low and high frequency components (LF, HF). Calculating SDNN and CVRR relied on the inter-beat intervals (IBIs) created in this software. The analysis included calculating HR delta and HRV delta. Calculation of these values included determining the difference between the heart rate or low frequency component of HRV for a workload condition and the baseline value. Analysis included computing ANOVAs to determine the cardiac measure(s) indicative of workload level .

ANALYSIS, RESULTS, AND DISCUSSION

The first measure investigated was the subjective workload (WWL), which is the weighted worked level calculated from the NASA TLX questionnaire. To determine if it was an accurate measure of the task load, six WWL scores were computed for each participant, one for each of the three workload levels, and each level was repeated, resulting in 78 samples. Analysis applied JMP software using the Residual Maximum Likelihood (REML) method to compute ANOVAs. Any value p-value less than 0.05 was assumed significant.

The results showed that the task load had a significant effect on WWL ($F(3,25)=31.386$, $p<=0.0001$). Performing a pair wise comparison with a Tukey's honestly significant difference (HSD) test shows that there is a significant difference between all task load levels except the two medium levels. Figure 4 shows workload level as a function of task load. As shown in this figure, the subjective estimate of workload (WWL) increased from 44.63 to 57.12 to 62.71 to 74.71 as the task load increased from the low through the high condition. The WWL appears stable since it was not significantly different between participants or runs.
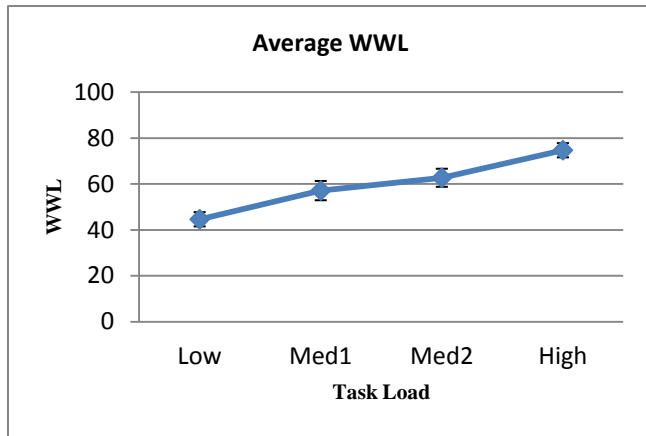
Figure 4: WWL by Task Load

The next measure investigated was the effect of task load on performance score. The JMP results show that both task load (F(3,24)=12.53, p<=0.0001) and run number (F(1,13)=12.14, p<=0.0039) had a significant effect on performance scores. A Tukey's HSD test showed that there was a significant difference in performance scores between low task load and the higher medium and high task loads, and that the performance for the high task load condition is significantly lower than the performance score for the lower medium and low task loads. There was no significant difference between the performance scores for the two medium levels of task load. The average performance score decreased from 0.65 to 0.57 to 0.55 to 0.48 as the task load increased from the low through the high condition. The average performance scores also increased 5 percent between runs. This may indicate that the participants would have benefitted from additional training.

The random effect of participant produced more variance for performance scores than it did for the WWL scores. This makes sense, as some participants will be more inclined to performing these types of tasks, while others may struggle greatly even with low task loads. These differences produce large variances in performance scores between participants making it difficult to point to a specific performance score as being indicative of high mental workload for a participant.

Heart rate measure analysis included applying the previously stated method for conducting ANOVAs. Table 1 summarizes the results of the analysis. Heart rate was significantly affected by task load (F(3,24)=6.34, p<=0.0026); however, it was only significant between the low and high task loads, which indicates that heart rate may only be able to discriminate between task loads with more extreme differences.

Performing the analysis on HRV (LF) yielded significant effects for task load (F(3,23.43)=6.98, p<=0.0016) and run number (F(1,12.76)=5.2951, p<=0.0389). The HRV (LF) value was highest at the low task load and lowest at the highest task load, which was expected. However, the two medium levels were neither significantly different from each other nor from the low or high levels. This may be because the medium levels have half the number of data points and

because participants, who introduce a high level of variance to the ANOVA model, only completed one of the medium levels.

Table 1: Heart Rate Measures with Significant Effects

HR Measure	Significant Effects	F Statistic	P-value
HF	Task Load	F(3,22.6)=3.8279	0.0235
	Run	F(1,15.77)=6.988	0.0179
	Task Load x Run	F(3,28.58)=3.3093	0.0341
SDNN	Task Load	F(3,23.39)=3.53	0.0305
CVRR	None	Not significant	Not significant
HRV Delta	Task Load	F(3,23.4)=3.72	0.0254
HR Delta	Task Load	F(3,23.7)=6.298	0.0027

Task load had a significant effect on the HF component of HRV, the SDNN, the HRV delta, and the HR delta. Additionally, the run number and the interaction of run number with task load had a significant effect on HRV (HF). Each of these measures showed trends similar to those seen for HR and HRV (LF). The CVRR measure did not show significant differences based on any of the fixed effects.

Analysis of the heart rate measures reveals some interesting trends. First all measures, except for CVRR, were significantly different for the low and high task loads, but were not significantly different with regard to the medium task loads. The responses for these task loads had significant variance. For some participants they followed the expected trends (i.e. HRV decreased as task load increased and HR increased with task load); however, for some participants the heart rate measures for medium task load did the opposite.

To determine which measures would be most appropriate for predicting a participant's perceived workload it was decided to use the Weighted Workload Level as the response since it represents the participant's perceived workload. The first step was to look at the correlations and partial correlations (the correlation with respect to all other correlations) between each measure and determine which attributes correlated most strongly with WWL. Analysis revealed a higher correlation of WWL with the performance than any of the heart rate measures. The partial correlations showed that HR delta from baseline showed the next highest correlation.

By building a linear regression model with WWL as the response and performance score and HR delta as the predictor variables the resulting R^2 value is 0.21 and the R^2 adjusted is 0.189, meaning that at most these two variables account for 18.9 percent of the variation in WWL. Exploration included applying stepwise regression to search for an improved model. The resulting model added HF as a predictor variable within the model, yielding an Adjusted R^2 value of 0.21.

Because the HR delta was chosen for each model, the data was also investigated with only the participants who had a more controlled baseline reading (participants 8-13). The stepwise regression of the model included WWL as the response variable and performance score, HR delta, HRV delta, and HF as predictor variables. The resulting model had a R^2 value of 0.61 and an Adjusted R^2 value of 0.55, which is a significant increase compared to the previous model, indicating the importance of a controlled baseline.

Discussion

The analysis included applying ANOVAs to explore the relationship between changes in task load and multiple heart rate measures, performance scores, and subjective workload scores. Changes in task load resulted in significant changes for each of these measures with the exception of CVRR. Regression models were also constructed to predict subjective workload levels, and by using the data from the subjects with more accurately controlled baseline data, an Adjusted R^2 value of 0.55 was achieved. The findings show promise in applying performance scores and various heart rate measures to estimate subjective workload and trigger automation in systems.

While most of the results were consistent with the background literature, a couple differences were apparent. In two studies, a significant correlation existed between CVRR and mental workload (Verwey & Veltman, 1996; Kawakita, Itoh, & Oguri, 2010). However, the current study found no significant effects for this measure. An interesting note is that each of the previous studies involved driving tasks, which may have a different effect on the various measures. Another interesting result pertained to the HRV (LF) measure. While this research indicated that it was sensitive to high and low task load differences, it was no more sensitive than other measures as in other studies (de Waard, 1996).

One interesting finding for this research came from the HR delta and HRV delta results. Both of these measures showed results similar to the other heart rate measures when performing the ANOVAs. However, when investigating potential regression models to predict WWL, these measures had a more significant effect than the other heart rate measures. Because the regression model is applied across participants, a delta measurement may provide a more accurate method to trigger adaptive automation.

CONCLUSION

This research investigated the concept of using various performance and heart rate measures to estimate workload with the goal of triggering adaptive automation. Research has shown that adaptive automation can increase overall performance, and that certain heart rate measures are indicative of an operator's mental workload. A human-participants experiment was developed to better understand how various heart rate measures and performance measures could be used in adaptive automation. The results suggested that certain measures can be used to distinguish between task loads, and that a combination of performance and cardiac measures can be used to model an operator's perceived mental workload.

REFERENCES

Averty, P., Athenes, S., Collet, C., & Dittmar, A. (2002). Evaluating a New Index of Mental Workload in Real ATC Situation Using Psychophysiological Measures. *IEEE* .

Booher, H. R. (2003). *Handbook of Human Systems Integration.* Hoboken, NJ: John Wiley and Sons, Inc.

de Waard, D. (1996). *The Measurement of Drivers' Mental Workload.* Haren, The Netherlands: The Traffic Research Centre VSC, University of Groningen.

Fedorovskaya, E.A., Miller, P., Prabhu, G., Horwitx, C., Matraszek, M. Parks, P. Blazey, R. and Endrikhovski, S. (2001) Affective imaging: Psychological and physiological reactions to individually chosen images, (pp. 524-532). SPIE 4299, San Jose, CA.

Kamshilin, A. A., Miridonov, S., Teplov, V., Saarenheimo, R., & Nippolainen, E. (2011, April 1). Photoplethysmographic imaging of high spatial resolution. *BIOMEDICAL OPTICS EXPRESS* .

Kawakita, E., Itoh, M., & Oguri, K. (2010). Estimation of Driver's Mental Workload Using Visual Information and Heart Rate Variability. *13th International IEEE*, (pp. 765-769). Portugal.

Miller, W. D. (2010). *The U.S. Air Force-Developed Adaptation of the Multi-Attribute Task Battery for the Assessment of Human Operator Workload and Strategic Behavior.* WPAFB: AFRL.

Novak, D., Mihelj, M., & Munih, M. (2010). Psychophysiological responses to different levels of cognitive and physical workload in haptic i nteraction. *Robotica* , 1-8.

Parasuraman, R. (2003). Adaptive Automation Matched to Human Mental Workload. In G. R. Hockey, *Operator Functional States* (pp. 177-193). IOS Press.

Parasuraman, R., & Byrne, E. (2003). Automation and Human Performance in Aviation. In P. S. Tsang, & M. A. Vidulich, *Principles and Practice of Aviation Psychology* (pp. 311-348). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Parasuraman, R., & Hancock, P. A. (2008). Mitigating the Adverse Effects of Workload, Stress, and Fatigue with Adaptive Automation. In P. A.

Parasuraman, R., Cosenzo, K. A., & De Visser, E. (2009). Adaptive Automation for Human Supervision of Multiple Uninhabited Vehicles: Effects on Change Detection, Situation Awareness, and Mental Workload. *Military Psychology* , 270-297.

Stanton, N. A., Salmon, P. M., Walker, G. H., Baber, C., & Jenkins, D. P. (2005). *Human Factors Methods: A Practical Guide for Engineering and Design.* Burlington, VT: Ashgate Publishing Company.

Tattersall, A. J., & Fairclough, S. H. (2003). Adaptive Automation and Modes of Control. In *Operator Functional State* (pp. 238-246). Liverpool John Moores University, UK: IOS Press.

Tsang, P. S., & Vidulich, M. A. (2003). *Principles and Practice of Aviation Psychology.* Mahwah, NJ: Lawrence Erlbaum Associates.

USAF. (2009). *United states air force unmanned aircraft systems flight plan 2009-2047.* Washington DC: Headquarters: United States Air Force.

Verwey, W. B., & Veltman, H. A. (1996). Detecting Short Period of Elevated Workload: A Comparison of Nine Workload Assessment Techniques. *Journal of Experimental Psychology: Applied* , 270-284.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering Psychology and Human Performance* . Upper Saddle River, NJ: Prentice-Hall Inc.

Appendix E

Published Article: Jeroski, J., Miller, M.E., Langhals, B. and Tripp, L. (in press). Impact of Vigilance Decrement upon Physiology Measures. Proceedings of the Industrial and Systems Engineering Research Conference, Montreal, CA. (Awarded Best Paper - Human Factors).

Impact of Vigilance Decrement upon Physiology Measures

Justine Jeroski, Michael E. Miller and Brent Langhals
Air Force Institute of Technology
Wright Patterson AFB, OH 45433

Lloyd Tripp
Air Force Research Laboratory
Wright Patterson AFB, OH 45433

Abstract

Despite a long history of vigilance research, the relationship between the vigilance decrement and a broad range of physiology measures has not been fully documented. In an attempt to address this gap, an experiment was designed in which participants detected critical signals displayed at random during a 20-minute simulated air traffic control vigilance task. In addition to collecting performance data, cerebral oximetry and electrocardiography were utilized to collect a range of physiological signals from participants including heart rate, heart rate variability, and cerebral oxygenation levels in the right and left frontal areas of the brain. The physiology data when correlated with the decrement indicated by the performance data demonstrated a potential relationship between these measures. This research has implications for using physiology measures to determine the onset of human vigilance decrement to institute compensatory measures.

Keywords

Vigilance decrement, Cerebral Oximetry, Electrocardiography, Performance

1. Introduction

Throughout history people have evolved technology, bringing about more complex and advanced systems that can perform tasks and operations faster and more accurately than before. As a result, tasks that once required physical and cognitive effort to perform can now be performed by systems through automation [1, 2]. With these advancements the role of the individual has changed from one of active involvement to one of passive supervision [3, 4]. As these systems are potentially fallible, human operators are left to scan information created by these new systems, to monitor their status and act only when an infrequent, but critical event, such as a system failure or emergency, arises [3]. Thus vigilance, or the ability to maintain attention and alertness over prolonged periods of time while monitoring for rare stimuli among frequently occurring stimuli, is required [2, 5-8].

Whereas it is relatively easy for people to be briefly attentive to a series of predictable events, maintaining attention to unpredictable events over a long period of time is difficult, especially when the events also have a low probability of occurrence. This decline in performance over time is known as the vigilance decrement [5, 7, 9]. The vigilance decrement typically appears within the first 15 minutes of watch, but depending on the nature of the task and the demand required, it could appear as rapidly as 5 minutes [5, 10]. Additionally, the vigilance decrement affects both novice and expert users [3]. As a result, automated human-machine systems, which have obvious benefits in the work place, have created problems related to over-reliance and waning vigilance [11].

The presence of the vigilance decrement has been well documented; however, the underlying cause of this performance decline is subject to debate. Currently, there are two competing theories, the mindless, boredom or under-load theory, and the resource, mental fatigue or over-load theory. Initially the under-load theory was developed with the belief that the vigilance decrement was caused by a decline in arousal or attention to a monotonous task. This theory hypothesizes that vigilance task participants' minds would wander, leading them to think thoughts unrelated to the task causing distraction, which eventually would lead to lack of awareness of the critical signals and decreasing detection rate [3, 7, 12, 13]. More recently a new theory has proposed that vigilance

tasks are difficult, stressful, and impose substantial demands on the information-processing resources of the individual. The resource theory proposes that the vigilance decrement is instead due to a decline in available attentional resources. Observers during vigilance tasks are required to make many decisions under uncertain conditions without rest. The continuous nature of this task does not allow for time to replenish resources. As a result mental resources become depleted over time, reducing the critical signal detection rates [3, 7, 9, 13].

As the concepts of vigilance and the vigilance decrement have developed and evolved, interest in research on this topic has gained momentum. This research is further motivated by the existence of vigilance tasks in a variety of military, industrial and medical settings, specifically the areas of air traffic control, cockpit monitoring, industrial process/quality control, airport baggage inspection, long-distance driving, robotic manufacturing, and cytological screening [3, 4]. Starting in the early twentieth century, individuals such as Henry Head who first described vigilance in brain injured patients and Norman Mackworth who studied vigilance during World War II using the famous "Clock Test" had began to quantify the source of decreased performance during vigilance tasks [3, 7, 10, 13]. The field expanded as researchers investigated human attention during vigilance tasks and human performance with a variety of different systems [3]. Studies of vigilance emerged with a variety of different characteristics including, multiple difficulty levels, event rates, task durations, stimuli types, and discrimination types, each with a unique experiment, but a similar goal [5].

Recently the field has begun to employ physiology measures to gain new perspectives into the causes of the vigilance decrement. While many physiological measures of the vigilance decrement have been examined and analyzed, there has been no agreement on a preferred method that clearly identifies loss of vigilance in every individual and in a variety of situations. Previous studies of vigilance using physiology measures have included one or more techniques to explore signals from the parts of the body such as the brain, eyes, heart, and skin [2, 4, 6, 14-16]. The focus of this study was to employ two of the most prevalent and readily available techniques, cerebral oximetry and electrocardiography (ECG) to further the understanding of the relationship between signals from the human body and the vigilance decrement.

Many studies have used cerebral blood oxygen saturation (rSO_2) using cerebral oximetry to quantify a vigilance decrement. There have been reasonably consistent results reported even with a variety of possible experimental factors. Most studies have found that a decline in performance is paralleled with an increase in oxygen saturation [2, 17]; however, others have found no significant changes in oxygen saturation values with time-on-task [6, 16]. Greater activity in the right over the left cerebral hemisphere has been reported for easier vigilance tasks, while a bilateral activation in oxygen saturation across the two hemispheres have been reported for more difficult vigilance tasks [2, 6, 15, 16].

An electrocardiogram (ECG) is a recording of the electrical activity of the heart over a period of time. It can be used to measures an individual's heart rate and heart rate variability and has been used in numerous studies to assess mental workload. Specifically heart rate variability (HRV) has been shown to have an inverse correlation with mental workload [12, 18]. Monitoring heart rate data during a vigilance task could therefore help to understand the amount of mental effort and whether the vigilance decrement is due to low mental effort (mindless theory) or instead due to a high mental effort (resource theory) [12].

The purpose of this study is to detect and quantify the vigilance decrement using well known methods as to establish control data to compare for further changes in experimental procedure. Additionally, this study plans to make direct comparisons between the effects of a variety of physiology measures with changes in performance all recorded during the same vigilance task.

2. Method

2.1 Participants
The participants enrolled in this study were volunteers from military and civilian employees of Wright Patterson Air Force Base. All participants had normal or corrected-to-normal visual acuity as verified using a SLOAN Multiple Group Near Vision Testing Card from Precision Vision and normal depth perception as determined using a TNO test for stereoscopic vision from Laméris Instrumenten b.v. Of the 33 participants, 21 were male and 12 were female. They ranged in age from 22 to 40 years with a mean of 28 years (SD = 5.1). And they included a representative

sample of individuals, with 29 right-handed, 3 left-handed, and 1 neutral-handed participants, based upon the results of the Edinburgh Handedness Survey [19].

2.2 Apparatus and Equipment.

A Metronaps EnergyPod, shown in Figure 1, allowed for containment of the participant during the experiment. The pod is 212.19 cm long, 145.73 cm tall, and the dome of the pod is 121.91 cm wide. The shield built into the pod functioned to prevent participants from becoming distracted by outside stimuli. The pod also permitted participants to sit in a relaxed and comfortable position throughout the experiment.

Figure 1: The pad and blanket situated inside the pod while the shield is open and closed.

Participants wore two sensors connected to a Somanetics Invos Cerebral Oximeter 5100B as shown in Figure 2. This system uses near-infrared spectroscopy technology to continuously and noninvasively measure blood oxygen saturation levels in the frontal areas of the left and right hemispheres of the brain. The near-infrared sensors were positioned and secured to the forehead of each participant using an adjustable headband. Care was taken to avoid the sinus cavities and any hair that might interfere with the signal. Both sensors were cleaned and tested for an effective reading between all participants.

Figure 1: Somanetics Invos Cerebral Oximeter 5100B (left) and participant wearing adjustable headband and EOG electrodes (right).

A BIOPAC© 150 was used to perform electrocardiography (ECG). Electrodes worn on the participant's chests were attached to the BIOPAC hardware system containing an ECG amplifier for measuring the electrical signals associated with the beat of the human heart. The BIOPAC© system with electrodes are shown in Figure 3. The BIOPAC© hardware system fed the signals into data acquisition software, AcqKnowledge®, where they were recorded and saved.

Figure 2: BIOPAC$^{©}$ hardware system showing placement of ECG electrodes.

2.3 Procedure

All participants performed the task individually in a quiet, windowless laboratory. Each participant was given an informed consent to read and sign prior to the experimental session. After completion of the consent form, the participant's visual acuity and depth perception were evaluated. Additionally their handedness was determined as right, left, or neutral using the Edinburgh Handedness Inventory. This data was recorded along with the participant's age and gender.

An in-depth description of the simulated air traffic control vigilance task was read and examples of critical and neutral signals were shown to each participant. The task involved a random presentation on a computer screen of three concentric circles with four arrows between the two outermost circles as shown in Figure 4. Participants viewed approximately 30 displays/min and each display remained on the screen for 1 second. The configuration of the four arrows between the two outer most circles changed each time the display was updated. Displays showing arrows aligned in a potential collision path were considered critical events and warranted an overt response from the participants. Participants indicated a critical event by pressing a finger mouse held in their dominant hand. Displays that showed arrows aligned in a non-collision path or safe path were considered neutral events and required no overt response from the participants. Examples of possible critical event displays and neutral event displays are shown in Figure 4. The software package was programmed to display 10 critical events randomly within each 10 minute portion of the 20-minute vigilance task, providing signal probability per period of 0.133 percent.
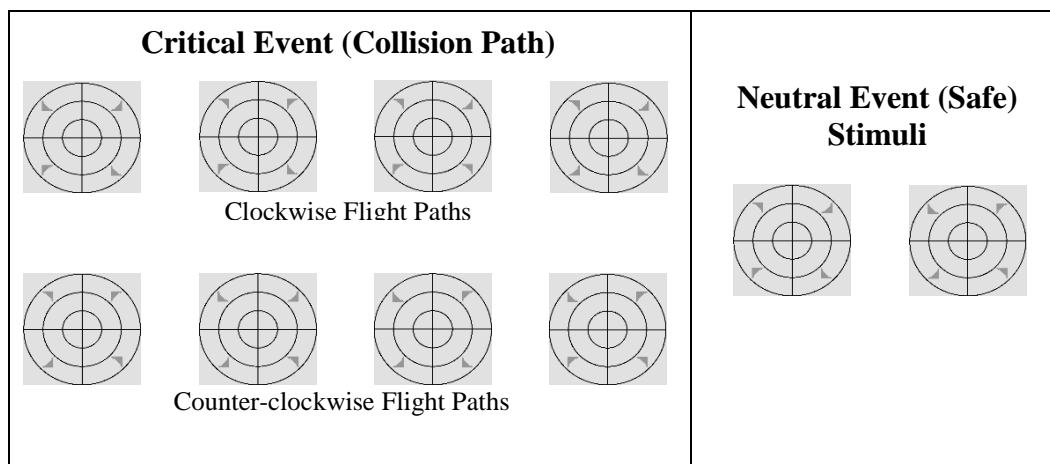

Figure 3: Possible critical event and neutral event displays from the simulated air traffic control vigilance task.

Prior to the beginning of the practice and test sessions, the participants were asked to sit in the pod and the cerebral oximeter and ECG sensors were fit to the appropriate areas of their face and body. All sensors were connected to their corresponding hardware and the resulting signals were tested to determine correct setup and sensor placement. Cerebral oximetry data was sampled at 0.2 Hz and recorded in a text file by HyperTerminal communication

software. The ECG data was sampled at a frequency of 1000 Hz and recorded by the BIOPAC AcqKnowledge® software.   A thermoelectric blanket was placed over the body of the participant and used to modulate the temperature of the participants in a subsequent set of trials, not discussed in this paper. A 48.26 cm (19 inch) computer monitor, used for displaying the vigilance task, was then mounted at eye-level inside the pod, approximately 60 cm from the participant.

Before the vigilance task test session, participants were given a 5-minute practice session, after which their hit and false alarm rate were calculated and feedback was given. Once the practice period was completed, participants began the 20-minute simulated air traffic control task. Custom software recorded the participant's response to every display throughout the vigilance task. Additionally, the communicator program and BIOPAC AcqKnowledge® software continued to record all physiology data. The data for this study was terminated at the 20-minute mark and the results were compiled and backed up to a secure computer. An additional period was then completed where the participants continued the vigilance task during which the temperature of the thermoelectric pad and blanket was modulated. However, the data associated with the temperature change is not reported in this paper. Finally, all the physiology sensors were removed from the participants and they were permitted to exit the pod.

2.4 Data Analysis

This experiment was conducted to demonstrate the vigilance decrement over a 20-mintue trial of a vigilance task and to look for trends in the physiology data that mirrored the decline in performance. As more workplaces rely on automated human-machine systems, sustained vigilance and the methods that detect vigilance has become increasingly important. Therefore the correlation of physiology data with the vigilance decrement provides potential methods for automated detection of vigilance loss.

A well-known and previously used vigilance task [4, 17], a simulated air-traffic control task, was chosen for the participants to complete. For each event, three types of responses were recorded during the experiment: a hit, which is a response during a critical signal; a miss, which is no response during critical signal; and a false alarm, which is a response during a neutral signal. All other events were neutral signals with no responses. From this data, the percent correct detection or hit rate and the false alarm rate were looked at for each participant individually and for the collective pool of participants.

To effectively compare the physiology measures across participants, a baseline period needed to be determined for each measure. Previous studies have used a 5-minute period prior to the task session to calculate a baseline for cerebral activity [2, 6, 16, 17, 20]. However, this baseline may not be available in the operational domain, thus a more easily used baseline was determined. Cerebral oxygen saturation values recorded during the first 2 minutes of the vigilance task were averaged and used as a baseline. A percent change from the baseline was calculated for each recorded value using the following equation, %Change = ((Recorded Value – Baseline) / Baseline) * 100. An average percent change for both the left and right hemispheres was calculated for each of the remaining 2 minute periods.

Heart rate (HR) and heart rate variability (HRV) were both also analyzed as a percent change from a baseline. Heart rate was calculated from the electrocardiogram (ECG) data by identifying the location and counting the number of the R wave peaks present during normal cardiac function. The heart rate baseline for the ECG was calculated by averaging the number of R waves over the first 2 minutes of the first vigilance task period. A percent change from the baseline was calculated using the average HR for each of the remaining 2 minute periods with the following equation, %Change = ((Average for 2 Minute Period – Baseline) / Baseline) * 100. One component of heart rate variability was determined by calculating the R – R Interval (RRI), or the time between each R wave. The standard deviation of these RRIs, known as SDNN, was calculated and then divided by the average RRI value to determine the coefficient of variation of R-R (CVRR). The same baseline technique and CVRR calculation was used to determine the values for each of the remaining 2 minute periods.

Repeated measures ANOVAs were applied to determine the statistical differences in performance scores, cerebral oximetry values, and heart rate measures between the baseline period and each of the additional nine 2 minute periods. Bonferroni post-hoc tests were conducted if an overall significant difference in means was found to determine where those differences occurred. A Greenhouse-Geisser correction was applied when the assumption of sphericity was violated. The level of significance having a probability of 0.05 was established a priori. All calculations were completed using MATLAB R2012a and SPSS Statistics 18.0.

3. Results

3.1 Performance

Performance was assessed in terms of the percentage of correct detections (or hit rate) and false alarms. A repeated measures ANOVA indicated a statistically significant effect for percent correct detections over time, $F(9, 288) = 2.133$, $p = 0.027$. This data is displayed in Figure 5.
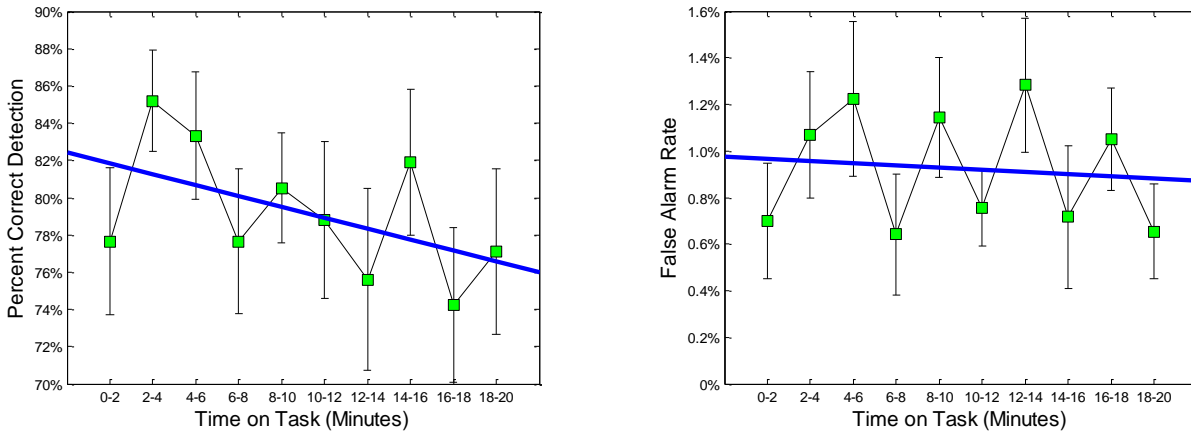

Figure 4: Mean percent correct detection (left) and mean false alarm rate (right) over periods of watch on the vigilance task. Bold lines indicated linear trends and error bars indicate plus and minus standard error of the mean.

A repeated measure ANOVA determined that there was no significant effect of percentage of false alarms (or false alarm rate) over time. A trend line showed a relatively stable rate over time and throughout the task, the number of false alarms was very low. This data is also displayed in Figure 5.

3.2 Cerebral Oximetry

A repeated measures ANOVA with a Greenhouse-Geisser correction indicated a statistically significant effect of cerebral oximetry values over time, $F(3.989, 123.645) = 3.073$, $p = 0.019$ and a significant effect between percent rSO_2 change values for right and left hemispheres, $F(1, 31) = 4.539$, $p = 0.041$. A trend line was added to the data to show the overall decline over time. This data is displayed in Figure 6.

3.3 Heart Rate

Heart rate measures from the ECG data included heart rate (HR) and heart rate variability (HRV). A repeated measures ANOVA determined that a significant decrease was present between time periods for percent HR change; $F(8, 256) = 7.093$, $p = 0.0000$. Post hoc tests using the Bonferroni correction revealed that there was a significant decline from period 2 to periods 8 ($p = 0.007$) and 9 ($p = 0.008$) and period 3 to periods 8 ($p = 0.007$) and 9 ($p = 0.007$). The data is displayed in Figure 7.

A repeated measures ANOVA with a Greenhouse-Geisser correction indicated there was a significant increase between time periods for percent HRV (CVRR) change; $F(5.724, 183.177) = 3.407$, $p = 0.004$. Post hoc tests using the Bonferroni correction revealed that there was a significant increases from period 2 to periods 5 ($p = 0.044$) and 8 ($p = 0.029$). The data is also displayed in Figure 7.
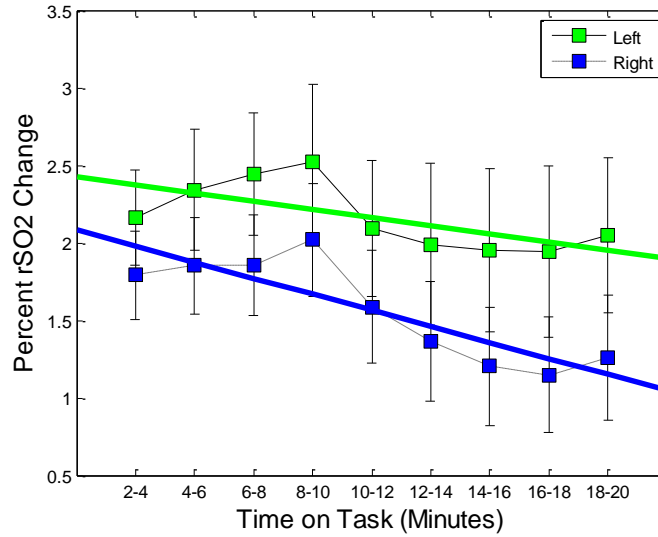
Figure 5: Mean oxygen saturation scores for the left and right hemispheres over the period of watch. Oxygen saturation scores are based upon percent change relative to baseline. Bold lines indicated linear trends and error bars indicate plus and minus one standard error of the mean.
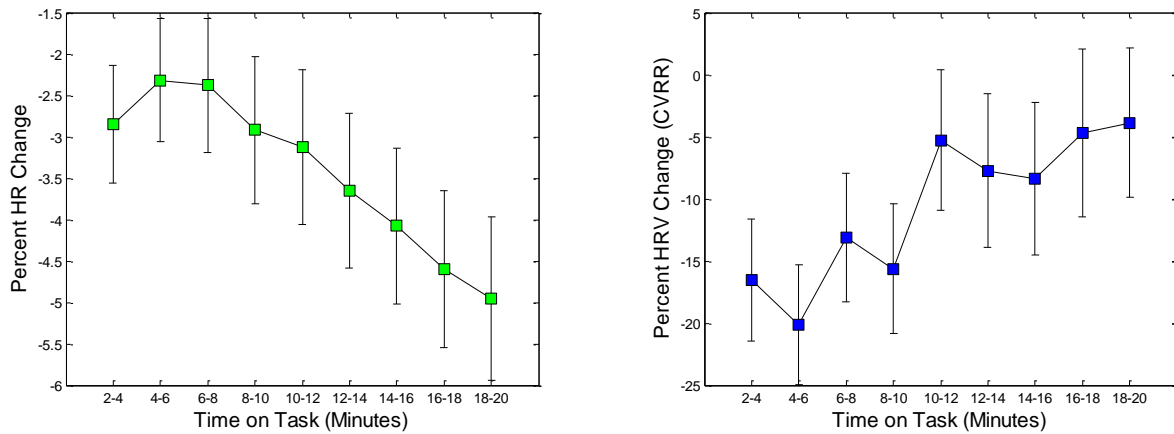

Figure 6: Mean percent heart rate (HR) change and percent heart rate variability (HRV) over the period of watch. Heart rate measures are based upon percent change relative to baseline. Error bars indicate plus and minus one standard error of the mean.

3.4 Correlations with Vigilance

A Pearson product-moment correlation was run to determine if there was a relationship between vigilance, or percent correct detections and any of the physiology measures. The component of HRV, coefficient of variation of R-R (CVRR) showed a significant result. All other measures were not significantly correlated with percent correct detections (hit rate). Additionally, all the physiology measures were significantly correlated with each other, as expected, considering all were measures related to blood flow. The correlation table is displayed in Table 1.

Table 1: Correlation Table. Correlation values that are * are significant at the 0.05 level (2 – tailed).

	Hit Rate	Left rSO$_2$	Right rSO$_2$	Heart Rate	CVRR
Hit Rate	1				
Left rSO$_2$	.341	1			
Right rSO$_2$	.538	0.928*	1		
Heart Rate	.544	0.760*	.895*	1	
CVRR	-.740*	-.737*	-.831*	-.823*	1

4. Discussion

Given that the advancement and increased complexity of systems has changed the role of the individual from active involvement to passive supervision within select fields; the need to maintain attention or vigilance over prolonged periods of time has become critical for success [1-3]. This requirement has found its way into the jobs of air traffic controllers, unmanned aerial system operators, TSA inspectors, and medical screening technicians [3, 4]. Therefore, determining a method to monitor the sustained attention of these individuals and intervene when a vigilance decrement occurs could improve overall task performance and reduce errors.

The fact that cerebral oxygen saturation levels increased during the first 10 minutes of the vigilance task (periods 1 to 4), suggests that increased processing demands on the brain called for increased oxygen supply to the tissue. High demand cognitive function activities cause more neurons to fire in the brain which burns glucose. A by-product of this activity is carbon dioxide (CO_2). An increase in CO_2 leads to vasodilatation, which results in increased blood flow to the region to remove the unwanted by-products [17, 21], and therefore an increase in available oxygen levels. As observed, from the 10 minute point, cerebral oxygen saturation levels then begin to decrease. This phenomenon could result from decreasing demand for blood flow, which occurs as available resources reach their maximum capacity. As participants' resources are depleted, their ability to perform the vigilance task diminishes. Less demand on cognitive function then results in the opposite effect, which causes a decrease of blood flow and less oxygen present [17]. This result fits well with the resource model of vigilance and shows cerebral oxygen saturation provides an index of utilization of information-processing resources during sustained attention [22].

Greater percent rSO$_2$ changes were seen in the left over the right hemispheres; however, there was no significant interaction between period and hemisphere, signifying bilateral activation. Difficult to discriminate targets and quick display rates [5], along with a lower percent correct detection rate (77 to 81 percent), would qualify this task as difficult. Therefore, it can be expected that the vigilance task initially placed increased processing demands on the brain in line with other difficult vigilance tasks. Previous research has suggested that increasing task difficulty induces a processing strategy change, from unilateral toward bilateral activation [2, 6, 15, 16].

The ECG calculated measures also showed results in line with resource theory. Percent heart rate (HR) and heart rate variability (HRV) change both changed significantly over periods of watch. HR results showed a negative percent change initially from the baseline and then decreases further with time on task. HR was positively correlated with vigilance; however not significantly. These findings are in line with previous studies that have positive correlations between heart rate and performance scores [23]; however, others have reported negative correlations or no significant effects either way [4, 24]. The component of heart rate variability (CVRR) in contrast, showed a statistically significant negative correlation with vigilance performance. This agrees with previous research, that is, HRV has been shown to have inverse correlation with mental workload [12, 18]. Participants were

engaged initially in correctly detecting all critical signals; however, as time progressed their physiology signals changed in relation to the decline in performance.

5. Conclusion

The results from this 20-minute vigilance study can add further information to the field of vigilance research with the goal of being able to identify a vigilance decrement in individuals to enable counter measures to be deployed. Ideally, the deployment of such countermeasures will permit the user to achieve greater success during tasks and activities requiring sustained attention. A variety of characteristics such as duration length, type of task, event rates, and modality could have had an effect on the results of this study; however, the physiology findings were similar to previous research and pointed to the resource model theory of vigilance. Percent correct detections or hit rate decreased over time and a significant decrement was determined at the 17-18 minute mark, in range of when a decrement is usually seen. Cerebral oximetry data showed a change from increasing to decreasing percent rSO_2 change at this point in the experiment. Heart rate measures also changed significantly near this point in the experiment. This data can add to the field of vigilance research and provide a controlled measure for this task and these physiology signals from which future research can be conducted.

Acknowledgements

References

1. Parasuraman, R., and Riley, V., 1997, "Humans and automation: Use, misuse, disuse, abuse," Human Factors: The Journal of the Human Factors and Ergonomics Society, 39(2), 230-253.
2. De Joux, N., Russell, P., and Helton, W., 2013, "A functional near-infrared spectroscopy study of sustained attention to local and global target features," Brain and cognition.
3. Warm, J., Parasuraman, R., and Matthews, G., 2008, "Vigilance requires hard mental work and is stressful," Human Factors: The Journal of the Human Factors and Ergonomics Society, 50(3), 433-441.
4. McIntire, L., Goodyear, C., Bridges, N., McKinley, R. A., Merritt, M., Griffin, K., and McIntire, J., 2011, "Eye-Tracking: An Alternative Vigilance Detector," Infoscitex Corp, Dayton, OH.
5. See, J., Howe, S., Warm, J., and Dember, W., 1995, "Meta-analysis of the sensitivity decrement in vigilance," Psychological Bulletin, 117(2), 230.
6. Helton, W., Hollander, T., Warm, J., Tripp, L., Parsons, K., Matthews, G., Dember, W., Parasuraman, R., and Hancock, P., 2007, "The abbreviated vigilance task and cerebral hemodynamics," Journal of Clinical and Experimental Neuropsychology, 29(5), 545-552.
7. Helton, W., and Warm, J., 2008, "Signal salience and the mindlessness theory of vigilance," Acta psychologica, 129(1), 18-25.
8. Stevenson, H., Russell, P., and Helton, W., 2011, "Search asymmetry, sustained attention, and response inhibition," Brain and cognition, 77(2), 215-222.
9. Temple, J., Warm, J., Dember, W., Jones, K., LaGrange, C., and Matthews, G., 2000, "The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task," Human Factors: The Journal of the Human Factors and Ergonomics Society, 42(2), 183-194.
10. Teichner, W., 1972, "Predicting Human Performance III: Detection of a Simple Visual Signal as a Function of Time of Watch," No. NMSU-ONR-TR-72-1, New Mexico State University Las Cruces Department of Psychology.
11. Molloy, R., and Parasuraman, R., 1996, "Monitoring an automated system for a single failure: Vigilance and task complexity effects," Human Factors: The Journal of the Human Factors and Ergonomics Society, 38(2), 311-322.
12. De Waard, D., 1996, "The measurement of drivers' mental workload," Groningen University, Traffic Research Center.
13. Helton, W., and Russell, P., 2011, "Working memory load and the vigilance decrement," Experimental brain research, 212(3), 429-437.
14. Oken, B., Salinsky, M., & Elsas, S., 2006, "Vigilance, alertness, or sustained attention: physiological basis and measurement," Clinical Neurophysiology, 117(9), 1885-1901.

15. Helton, W., Ossowski, U., and Malinen, S., 2013, "Post-disaster depression and vigilance: a functional near-infrared spectroscopy study," Experimental brain research, 1-6.

16. Helton, W., Warm, J., Tripp, L., Matthews, G., Parasuraman, R., and Hancock, P., 2010, "Cerebral lateralization of vigilance: a function of task difficulty," Neuropsychologia, 48(6), 1683-1688.

17. Funke, M., 2009, "A Comparison of Cerebral Hemovelocity and Blood Oxygen Saturation Levels During Vigilance Performance," Doctoral dissertation, University of Cincinnati.

18. Rowe, D., Sibert, J., and Irwin, D., 1998, "Heart rate variability: Indicator of user state as an aid to human-computer interaction," In Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press/Addison-Wesley Publishing Co., 480-487.

19. Oldfield, R., 1971, "The assessment and analysis of handedness: the Edinburgh inventory," Neuropsychologia, 9(1), 97-113.

20. Beam, C., 2002, "Effects of sensory and cognitive vigilance tasks on cerebral blood flow velocity," Doctoral dissertation, University of Cincinnati.

21. Bélanger, M., Allaman, I., and Magistretti, P., 2011, "Brain energy metabolism: focus on astrocyte-neuron metabolic cooperation," Cell metabolism, 14(6), 724-738.

22. Hitchcock, E., Warm, J., Matthews, G., Dember, W., Shear, P., Tripp, L., Mayleben, D., and Parasuraman, R., 2003, "Automation cueing modulates cerebral blood flow and vigilance in a simulated air traffic control task," Theoretical Issues in Ergonomics Science, 4(1-2), 89-112.

23. Eason, R., Beardshall, A., and Jaffee, S., 1965, "Performance and physiological indicants of activation in a vigilance situation," Perceptual and motor skills, 20(1), 3-13.

24. Dussault, C., Jouanin, J., Philippe, M., and Guezennec, C., 2005, "EEG and ECG changes during simulator operation reflect mental workload and vigilance," Aviation, space, and environmental medicine, 76(4), 344-351.

REPORT DOCUMENTATION PAGE

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to an penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
30 Apr 2014	Interim Report	April 2013-April 2014

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Workload-Adaptive Human Interface to Aid Robust Decision Making in Human-System Interface – Year 1 Report	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
	13RSL-AFIT1
Michael E. Miller, PhD	5e. TASK NUMBER
Gilbert Peterson, PhD	
LtCol Brent Langhals, PhD	5f. WORK UNIT NUMBER
Capt. Jason Bindewald	

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)	8. PERFORMING ORGANIZATION REPORT NUMBER
Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way WPAFB OH 45433-7765 DSN: 785-3636	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
James H. Lawton, PhD Program Officer, Cognition and Decision US Air Force Office of Scientific Research	AFOSR
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution Statement A. Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES
This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT
This report summarizes activity from the first year of a three-year project on aiding decision making in human-system interface. This report includes description of a new adaptive automation design model, allowing system designers the ability to visually and systematically evaluate the placement of adaptive automation within a system network. This model aids the designer to isolate the tasks, which require human decision making, ideally permitting tasks to be automated which do not require human decisions. A tool for evaluating adaptive automation, referred to as Space Navigator has been developed. This game has been designed using the model, has allowed us (and future researchers) to simplify data gathering from human participants. The resulting automation system will allow for research into how similarity and difference of actions between a human-machine team affect the overall performance of the system. The final experimental data will provide several areas of further research including trust in automation, training improvement, workload reduction (actual and perceived), and task load switching.

15. SUBJECT TERMS
Decision Making, Interface, Adaptive Automation, Design Tools

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Michael E. Miller
REPORT U	ABSTRACT U	c. THIS PAGE U	UU		19b. TELEPHONE NUMBER *(Include area code)* (937) 255-3636 x4651

Standard Form 298 (Rev: 8-98)
Prescribed by ANSI Std. Z39-18